604/17-2-77

Psychological Review

RICHARD L. SOLOMON, Editor
University of Pennsylvania

VOLUME 71, 1964

HELEN ORR
Managing Editor

ELIZABETH S. REED
Advertising Manager

RITA KOPANIC
Editorial Assistant

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA. 17604
AND 1200 SEVENTEENTH ST. N. W., WASHINGTON, D. C. 20036

Editorial Consultants

Robert P. Abelson
Norman H. Anderson
Justin M. Aronfreed
Elliott Aronson
John W. Atkinson
Albert Bandura
Neil R. Bartlett
William F. Battig
Frank A. Beach
Wesley C. Becker
D. E. Berlyne
James Bieri
Arnold M. Binder
Abraham H. Black
C. Alan Boneau
Lyle E. Bourne, Jr.
Jack W. Brehm
William E. Broen
Donald M. Borverman
Judson S. Brown
Jerome S. Bruner
F. Robert Brush
Robert R. Bush
J. Merrill Carlsmith
Raymond B. Cattell
Alphonse Chapanis
Irvin L. Child
Clyde H. Coombs
Judith L. Crooks
James Deese
James C. Diggory
James P. Egan
Charles W. Eriksen
Seymour Feshbach
Robert Galambos
Eugene Galanter
Norman Garmezy
Wendell R. Garner
Eleanor J. Gibson
James J. Gibson
Kenneth P. Goodrich
David A. Grant
David M. Green
J. P. Guilford
Norman Guttman
Eric C. Heineman
Winfred H. Hill
Davis H. Howes, III
J. McV. Hunt
Lloyd G. Humphreys
Lee M. Hurvich
Francis W. Irwin
George A. Kelly
Harold H. Kelley
Howard H. Kendler
Irwin Kremen
Gregory A. Kimble
John I. Lacey
Herschel W. Leibowitz
Seymour Levine
Richard A. Littman
Frank A. Logan
R. Duncan Luce
Paul E. Meehl
Samuel J. Messick
Howard Moltz
O. Hobart Mowrer
Conrad G. Mueller
Bennet B. Murdock, Jr.
Clyde E. Noble
Jum C. Nunnally
James Olds
Martin Orne
Dean Peabody
Leo J. Postman
Floyd Ratliff
L. Starling Reid
Frank Restle
Henry W. Riecken
Milton E. Rosenbaum
Mark R. Rosenzweig
Burton S. Rosner
George E. Ruff
Harris B. Savin
Pauline S. Sears
Robert R. Sears
John P. Seward
B. F. Skinner
Karl U. Smith
Kendon Smith
Arthur W. Staats
Eliot Stellar
Saul H. Sternberg
Joseph C. Stevens
S. Smith Stevens
Patrick Suppes
John A. Swets
Philip Teitelbaum
John W. Thibaut
Ledyard R Tucker
Endel Tulving
Benton J. Underwood
William S. Verplanck
Morris S. Viteles
Richard D. Walk
Hans Wallach
Nancy C. Waugh
Wilse B. Webb
Robert W. White
Ben James Winer
Julius Wishner

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS OF VOLUME 71

Aronfreed, Justin. **The Origin of Self-Criticism** 193

Berkowitz, Leonard. **Aggressive Cues in Aggressive Behavior and Hostility Catharsis** 104

Bilodeau, Edward A., and Levy, C. Michael. **Long-Term Memory as a Function of Retention Time and Other Conditions of Training and Recall** 27

Black, A. H., and Lang, W. M. **Cardiac Conditioning and Skeletal Responding in Curarized Dogs** 80

Boersma, Frederick J., DeJonge, James J., and Stellwagen, Walter R. **A Power Comparison of the *F* and *L* Tests—I** 505

Coan, Richard W. **Facts, Factors, and Artifacts: The Quest for Psychological Meaning** 123

DeJonge, James J. See Boersma, Frederick J.

Denenberg, Victor H. **Critical Periods, Stimulus Input, and Emotional Reactivity: A Theory of Infantile Stimulation** 335

Deutsch, J. A. **Neurophysiological Contrast Phenomena and Figural Aftereffects** 19

Dodwell, P. C. **A Coupling System for Coding and Learning in Shape Discrimination** 148

Dorfman, Donald D. See Zajonc, Robert B.

Feather, N. T. **A Structural Balance Model of Communication Effects** 291

Feshbach, Seymour. **The Function of Aggression and the Regulation of Aggressive Drive** 257

Flock, Howard R. **A Possible Optical Basis for Monocular Slant Perception** 380

Gellhorn, E. **Motion and Emotion: The Role of Proprioception in the Physiology and Pathology of the Emotions** 457

Green, David M. **Consistency of Auditory Detection Judgments** 392

Grice, G. Robert, and Hunter, John J. **Stimulus Intensity Effects Depend upon the Type of Experimental Design** 247

Hammond, Kenneth R., Hursch, Carolyn J., and Todd, Frederick J. **Analyzing the Components of Clinical Inference** 438

Hammond, Kenneth R. See Hursch, Carolyn J.

Hardy, Kenneth R. **An Appetitional Theory of Sexual Motivation** 1

Houston, John P. **Verbal Transfer and Interlist Similarities** 412

Hunter, John J. See Grice, G. Robert.

Hursch, Carolyn J., Hammond, Kenneth R., and Hursch, Jack L. **Some Methodological Considerations in Multiple-Cue Probability Studies** 42

Hursch, Carolyn J. See Hammond, Kenneth R.

Hursch, Jack L. See Hursch, Carolyn J.

Jenkins, James J., and Schuell, Hildred. **Further Work on Language Deficit in Aphasia** 87

Kimmel, H. D. **Further Analysis of GSR Conditioning: A Reply to Stewart, Stern, Winokur, and Fredman** 160

Landauer, Thomas K. **Two Hypotheses Concerning the Biochemical Basis of Memory** 167

Lang, W. M. See Black, A. H.

Leitenberg, H. See Levine, M.

Levine, M., Leitenberg, H., and Richter, M. **The Blank Trials Law: The Equivalence of Positive Reinforcement and Nonreinforcement** 94

Levy, C. Michael. See Bilodeau, Edward A.

Liu, In-mao. **A Theory of Classical Conditioning** 408

LOCKARD, ROBERT B. **A Method of Analysis and Classification of Repetitive Response Systems** 141
MARTIN, JOHN. See STANLEY, GORDON.
MILLER, HOWARD. See WILSON, WARNER R.
MISCHEL, T. **Personal Constructs, Rules, and the Logic of Clinical Activity** 180
NORMAN, DONALD A. **A Comparison of Data Obtained with Different False-Alarm Rates** 243
POSNER, MICHAEL I. **Information Reduction in the Analysis of Sequential Tasks** .. 491
RESTLE, FRANK. **A Cognitive Interpretation of Intensity Effects in Stimulus Generalization** 514
RICHTER, M. See LEVINE, M.
ROHRACHER, HUBERT. **Comments on "A Resonance Theory of 'Microvibrations'"** ... 524
SCHUELL, HILDRED. See JENKINS, JAMES J.
SILVERMAN, JULIAN. **The Problem of Attention in Research and Theory in Schizophrenia** 352
SINES, JACOB O. **Actuarial Methods as Appropriate Strategy for the Validation of Diagnostic Tests** 517
SLAMECKA, NORMAN J. **An Inquiry into the Doctrine of Remote Associations** 61
SMITH, KENDON. **Comment on the Paper by Black and Lang** 86
SMITH, KENDON. **Curare Drugs and Total Paralysis** 77
STANLEY, GORDON, AND MARTIN, JOHN. **How Sincere is the Dogmatist?** 331
STELLWAGEN, WALTER R. See BOERSMA, FREDERICK J.
TODD, FREDERICK J. See HAMMOND, KENNETH R.
TREISMAN, MICHEL. **Noise and Weber's Law: The Discrimination of Brightness and Other Dimensions** 314
TUCKER, LEDYARD R. **A Suggested Alternative Formulation in the Developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd** 528
TULVING, ENDEL. **Intratrial and Intertrial Retention: Notes towards a Theory of Free Recall Verbal Learning** 219
VALENSTEIN, ELLIOT S. **Problems of Measurement and Interpretation with Reinforcing Brain Stimulation** 415
WEIR, MORTON W. **Developmental Changes in Problem-Solving Strategies** 473
WILLIAMS, JAMES G. L. **A Resonance Theory of "Microvibrations": A Reply to Rohracher** 526
WILSON, WARNER R., AND MILLER, HOWARD. **A Note on the Inconclusiveness of Accepting the Null Hypothesis** 238
ZAJONC, ROBERT B., AND DORFMAN, DONALD D. **Perception, Drive, and Behavior Theory** 273

VOL. 71, NO. 1 JANUARY 1964

PSYCHOLOGICAL REVIEW

AN APPETITIONAL THEORY OF SEXUAL MOTIVATION

KENNETH R. HARDY
Brigham Young University

Following a brief review of evidence on the adequacy of biological explanations of sexual motivation, a cognitive-affective model of motivation is stated. The model is used to develop a conception of sexual motivation as an experientially developed appetite. 2 constitutional sources of affect are proposed, each of which is sufficient to reward behavior: the pleasure of genital stimulation or arousal and the complex excitement-relaxation of orgasm. The nature of childhood sexuality and the elaboration of sexual appetite in courtship are examined. After a discussion of approach-avoidance dynamics, problems of sexual control and compatibility are considered.

The purpose of this paper is to develop a theory of human sexual motivation which may more adequately account for *human* behavior than the conceptions now prevalent in the psychological and lay literature. Despite much direct and indirect evidence to the contrary, some form of biological drive theory is dominant today, couched in terms of drive reduction, homeostatic disequilibrium, endocrine secretion, or instinctual libido.

To be proposed is a conception of sexual appetite based upon psychological hedonism or affective change. This statement owes much to the thinking and research of F. A. Beach, P. T. Young, D. C. McClelland, Kurt Lewin, D. O. Hebb, and many others. However, no previous attempt has been made, as far as is known, to systematically pursue the implications of affective theory to the development of sexual appetite. Further, the present author must remain responsible for this development, as those mentioned may wish to disown the particular formulation proposed here.

INADEQUACY OF ORGANIC EXPLANATIONS

It is beyond the scope of this paper to give a comprehensive review of evidence regarding the adequacy of biological needs in accounting for sexual motivation. Nevertheless, a brief summary of several lines of investigation may be stated:

Phyletic Comparisons. Much of the experimental work on sex motivation has been done on various species of birds and rodents (Beach, 1951). This work demonstrates the importance of hormonal factors, particularly in the females of these species. The tendency to extrapolate from these data to humans should be strongly tempered by several considerations. In the first place, there are striking interspecies differences. For example, although testosterone propionate treatment accelerates the rate of maturation of sexual responses and produces greater than normal sexual drive and secretory activity of the accessory sexual organs in the rat, it does not do so in the guinea pig (Gerall, 1958). Also, studies of higher mammals, especially primates, show the relaxation of endocrine and genetic controls, with experience playing an increasingly prominent role as the human level is approached (Beach, 1947; Ford & Beach, 1951; Harlow, 1962; Nissen, 1953). Finally, when constitutional or hormonal interpretations of sexual motivation are tested directly upon humans, they fail to adequately account for the data, as noted in what is to follow.

Infantile and Childhood Sexuality. Cross-cultural data show the extensive elaboration of sexual behavior (masturbation, mutual masturbation, sexual games, attempted coitus, coitus itself) in cultures which are permissive, encouraging, or demanding of sexual activity prepuberally (Ford & Beach, 1951; Henry & Henry, 1953). In such societies, virtually no secrecy is associated with sex, and little or nothing is done to inhibit the sexual activity of the youngsters. Ford and Beach (1951) report that in a few cultures, regular heterosexual intercourse occurs from ages 8–10 onward; e.g., among the Chewa it is believed that youth will not mature unless intercourse begins early in life; similarly, the Lepcha believe that girls will be infertile unless they participate in coitus prepuberally.

The extent of sexual activity in childhood among the peoples mentioned above is a powerful demonstration of the hypothesis that sexual interest is not dependent upon the maturation of the sex glands.

Effects of Castration. When males are castrated postpuberally, the effects are highly variable, ranging from heightened libido to an abrupt cessation of sexual motivation (Hamilton, 1943; Tauber, 1940). Tauber, after reviewing the literature, concluded that the attitude of the patient toward the castration was a critical factor.

When females are castrated postpuberally, the effects again are variable. In Filler and Drezner's (1944) careful study of 41 cases, 36 experienced no libidinal change, 3 showed a drop in libido, and 2 showed an increase.

Data are scanty regarding the effects of prepuberal castration. Ford and Beach (1951) refer to a few cases where the male is able to voluntarily obtain an erection, copulate, and experience orgasm (of course without ejaculation); these may be atypical in our culture, however. Since prepuberal castration usually results in underdevelopment of the penis, it would bring about lessened effectiveness in copulation, as well as all of the psychological implications of masculine inadequacy that a small penis may connote to the patient.

Homosexuality. Kinsey (1941) convincingly maintains that hormonal interpretations are inadequate to account for the following facts: (*a*) homosexuality and heterosexuality are not mutually exclusive, but coexist in combinations of every degree of relative preponderance; (*b*) persons may go from heterosexual to homosexual relationships and vice versa in a matter of hours; or alternatively, there may occur months or years of an exclusively heterosexual or homosexual relationship, followed by a reversal of pattern; (*c*) thus far, there is an inability to distinguish between heterosexuals and homosexuals on the basis of hormonal secretion or constitutional typologies.

In a few societies (Siwan, Aranda, Keraki, Kiwai), sodomy is virtually universal, with those not participating considered peculiar. Commonly, these relationships begin at puberty, the boy taking the feminine role at first, later assuming the masculine role (Ford & Beach, 1951). Such relations coexist, of course, with heterosexual relations. In their study of American men, Kinsey, Pomeroy, and Martin (1948) report that 37% of their sample had had at least one homosexual experience resulting in orgasm, though only 4% were exclusively homosexual.

It is true that most American men settle into a pattern of heterosexuality, while a small minority establish various degrees of bisexual orientation (Kinsey et al., 1948). Some members of this minority establish definite homosexual preferences, with an accompanying way of life which includes a special jargon, mien, and social life among those similarly disposed (Cleckley, 1957; Cory, 1961). Whether one defines homosexuality as dealing with these extreme cases, or subsumes under this term a more widespread personality predisposition, the biological interpretation is equally tenuous.

Hermaphroditism. Data on individuals with confused or mistaken sex identity show variations in sexual desire, fantasy, and social role to be a function of the person's self-conception as to which sex he is (usually this is dependent upon parental conceptions), and not upon actual sex identity as determined by gonadal tissue (Ellis, 1945; Finesinger, Meigs, & Sulkowitch, 1942; Miles, 1942; Seward, 1946). In Ellis' exhaustive review of the literature, those who were true hermaphrodites (possessing both masculine and feminine gonadal tissue) were not significantly different from the pseudohermaphrodites in direction of sex libido. Of the total group studied intensively by Ellis, 80% had heterosexual libido, 11% were psychosexually immature, and 9% were homosexual or bisexual in libidinal interest, percentages which are probably fairly comparable to a "normal" group. Furthermore, the only bisexuals (those having heterosexual and homosexual motivation) came from the pseudohermaphrodites, the group least expected (biologically) to produce them. In virtually every case, the sex role and libidinal interest were those expected of the sex the individual was assumed to be while growing up, with little correspondence between sex role and libidinal interest on the one hand, and somatic characteristics on the other.

More recently, Money, Hampson, and Hampson (1955a, 1955b, 1957) and Money (1961) have reported results of comprehensive medical and psychological observations on 105 hermaphroditic patients. Their results strongly confirm Ellis' findings, but are even more striking. Only five of their patients had a gender role or libidinal orientation ambiguous or deviant from that expected for the sex to which they were assigned and in which they were reared. This was true despite the fact that these patients had one or more of the following physical variables of sex which were incongruous with the assigned sex: external genital morphology, internal accessory reproductive structures, hormonal sex and secondary sex characteristics, gonadal sex, and chromosomal sex. Especially striking were cases of the same medical diagnosis (e.g., hyperadrenocortical female, cryptorchid male hypospadiac), some of whom had been reared as masculine, others as feminine. Those raised as masculine were strikingly different from those raised as feminine in gender role and libidinal orientation, which followed their assigned sex in virtually every case.

Treatment of Sexual Problems. Attempts to treat impotence and frigidity by hormonal means have proved generally unsuccessful, when the effects of suggestion (the meaning of the treatment to the patient) are controlled (Carmichael, Noonan, & Kenyon, 1941; Creevy & Rea, 1940; Seward, 1946). Seward (1946, pp. 198–200) concludes her discussion of relevant literature by maintaining that in virtually all cases of frigidity and impotence, the difficulty is psychogenic rather than organic in character, and that the efficacy of psychotherapy as against hormonal therapy substantiates this diagnosis.

Periodicity of Sex Desire. Data from Davis (1929), Dickinson and Beam (1931), Ellis (1936), Hamilton (1929), Terman (1938), Tinklepaugh (1933), indicate two peaks of sexual desire in women, one occurring immediately prior to menstruation, the other occurring just following menstruation, with desire at a low ebb during the period of peak fertility. This is exactly opposite to expectation based upon hormonal fluctuations. Benedek and Rubenstein (1939, 1942), on the other hand, report maximal heterosexual drive among patients in psychoanalysis during the follicular phase of maximum fertility. These investigators inferred drive levels from reports of fantasy obtained during psychoanalytic interviews. Further research is needed to clarify the differences in results, whether due to populations studied, relative validity of measures used, or other factors. There does seem to be general agreement that periodic variations in sex drive are difficult for many if not most women to report. If such monthly fluctuations were of major consequence, it seems reasonable to assume that they would be more fully reflected in behavior or consciousness.

Although it is part of our folklore that, following sexual outlet, postpuberal males experience a steadily increasing sex drive which becomes more and more insistent, this belief is most probably false. In the words of Beach (1956),

> No genuine tissue or biological needs are generated by sexual abstinence. It used to be believed that prolonged sexual inactivity in adulthood resulted in the progressive accumulation of secretions within the accessory sex glands, and that nerve impulses from these distended receptacles gave rise to sexual urges. Modern evidence negates this hypothesis. . . . What is commonly confused with a primary drive associated with sexual deprivation is in actuality sexual appetite, and this has little or no relation to biological or physiological needs [p. 4].

It seems warranted to conclude that the overwhelming proportion of the variance in human sexual motivation and behavior is not explicable in terms of some biological need or tension, however conceived. An alternative view is explored in this paper, one which uses a series of notions based largely on cognitive expectancy theory, Lewinian field theory, and modern hedonic or affective theory.

PRINCIPLES OF MOTIVATION

The present theory of sexuality is based upon several principles common to motives generally, which will be stated prior to the specific discussion of sex motivation.

1. Motives are based upon learned expectations of an affective (hedonic) change (McClelland, 1951, p. 466; Young, 1959). These expectations are aroused by cues or stimuli which are associated with affective states due to some previous learning. Such learning may be based on the actual experience of the individual; however, it may also be vicariously based on affect laden fantasies (imaginal processes) aroused either by the person's own thought processes or through some form of social influence (e.g., reading a romantic novel). It is not implied that imaginal or vicarious experience is necessarily as potent a source of motivation as firsthand, physical experience, since the physical experience will ordinarily carry more affect, and may

have a richer context of cues (Principles 4 and 5 below). However, imagined or vicarious experiences may be highly motivating. This is witnessed by common fears of ghosts, dragons, bogeymen, haunted houses; behavior following a horror movie (locking windows, insomnia, nightmares). Johnson's (1945) report of the phantom anesthetist of Mattoon is a dramatic illustration.

2. For such an expectation to motivate behavior, the person must believe that some action (symbolic or overt) on his part helps or is required to bring about the desired affective change. A motive, then, is an expectation that the pursuit of a given course of action will contribute to, or result in, a change in affective state. The explicit requirement made here that instrumental behavior is a necessary part of a motivational expectation seems to go beyond the statement of affective theory by McClelland, Atkinson, Clark, and Lowell (1953).

An example may help to clarify this point. Suppose a small boy hears the bells of an ice cream truck coming down the street; his mother buys him an ice cream bar; and he thoroughly enjoys this tasty treat. There will then be a tendency for the bells on subsequent occasions to produce in him an expectation of pleasurable affect. However, the behavior he is motivated to employ is dependent upon the way in which the thoughts or actions he engages in are conceived as mediating between the cues and the affect, or serve as affect-producing cues themselves. Thus if the boy had to do nothing to receive the ice cream bar, he would merely engage in watchful waiting upon hearing the bells. If he had to coax his mother to purchase him a bar, however, the sound of the bells would instigate vigorous action on his part.

3. The confidence one places in a given affective expectation increases each time it is confirmed through having the affect follow or accompany the cues which arouse the expectation. Conversely, the lack of confirmation (infirmation) of an expectation tends to destroy belief in it as the predicted affective changes fail to follow the associated cues (Bruner, 1951). Furthermore, the extent to which the affect and stimuli are uniquely associated determines the degree to which the affect will be aroused by only such cues and the degree to which the cues will arouse only such affect. Continuing the example of the boy, if the occasion mentioned is the only time he hears such bells and has such an ice cream bar, then it can be said that these two things have been uniquely associated. If he hears such bells subsequently and enjoys this tasty treat each time, then he will develop great confidence (faith) that when the bells are heard a pleasurable eating experience will follow. The principle of uniqueness is similar to the Bruner-Postman concept of monopoly (Bruner, 1951).

4. There is some positive relationship between the amplitude of the affective change and the intensity of the resulting motive.

5. The likelihood of arousal of a motive is a function of the variety of cues associated with the affective change, together with the associative strength of those cues to the affect (Principle 3).

The last three principles parallel rather closely the analysis by McClelland et al. (1953, pp. 68–73) of factors affecting motive strength.

6. A family of expectations is a group of expectations which tend to arouse one another or to be aroused by a similar set of cues. Confirmation of one member of a family of expectations strengthens belief in other members of that family. Conversely, in-

firmation of one member reduces belief in other members of the set. The amount of generalized confirmation or infirmation is a function of the organization of relationships between the confirmed member and the other members of the family (see Asch, 1952, pp. 580–582; Rokeach, 1960, ch. 2).

These statements are cognitive equivalents for phenomena often phrased in terms of stimulus generalization, discrimination, and habit family hierarchy.

7. There is a degree of habituation in some affective areas in which repetition or continuance of an activity leads to a lesser affective change than that which occurred upon initial participation in the activity. For example, viewing the Grand Canyon of the Colorado or the Lower Falls of the Yellowstone is more thrilling the first time seen than later. Somehow the "edge of newness" is rubbed off, perhaps as a function of gratification of the cognitive need to know the environment, or of other habituating factors (see Dember & Earl, 1957).

8. Motives can be grossly categorized into three types: (*a*) *Approach,* in which the expectations are of a positive affective change; (*b*) *Avoidant,* in which the expectations are of a negative affective change; (*c*) *Ambivalent,* in which both positive and negative expectations are aroused by the same cues.

9. The tendency to act on the basis of some motive is a function of (*a*) the relative net affective strength of the alternatives aroused in a given situation; and (*b*) the subjective probability of attainment of the goals involved (see Atkinson, 1957). The net affective strength of a course of action is the algebraic sum of the positive and negative affectivity associated with it. When the difference in Factors *a* and *b* between two alternatives is great, other things equal, the decision is relatively easy; when the alternatives are about equal, decision is difficult, with vacillation and decision time increasing (Cartwright, 1940; Cartwright & Festinger, 1943).

10. Other things equal, the more immediate the affective change the more attractive or compelling is a given alternative as opposed to a change more distant; that is, as the psychological distance to a goal increases, the less motivating it is (Miller, 1951). This time-distance gradient is much steeper for children than adults, who live in a wider time perspective (Lewin, 1946). The immediacy or delay of gratification may be related to the subjective probability or certainty of attainment. For the hunter, a bird in the hand is indeed worth two in the bush because he may not get the birds in the bush.

11. The perceived permanence or transience of the anticipated affective state is relevant in assessing the net affect of a given alternative. Stealing may bring money for a temporary splurge, but may bring lengthy incarceration, distrust by friends, chronic twinges of conscience. Other things equal, affective states perceived as being relatively permanent will be more compelling than those relatively transient. Again, the differences between children and adults in time perspective should be noted. This dimension will be less important for children because of their relative inability to foresee long-term consequences (Lewin, 1948, p. 105).

12. Another pertinent problem in the analysis of choice behavior is the problem of mutual exclusiveness. Can a person take one alternative and also, perhaps later, another; can he have his cake and eat it too? Obviously, the perception of mutual exclusiveness may make a decision much more difficult than when both goals are seen as jointly attainable.

13. Closely related to the above is the finality or irrevocability of the decision. For example, if marriage is conceived of as a permanent bond between two people, the decision to marry a given person will be a more difficult one than if marriage is seen as an arrangement to be dissolved at the convenience of one or both of the partners.

With the foregoing tentative principles stated in general form, attention can now be directed specifically to sexual motivation.

AFFECTIVE FOUNDATIONS OF SEXUALITY

A theory which assumes that motives are learned in association with affective experience must, in order to be complete, describe those conditions which produce affective change in the first place. McClelland, recognizing this problem, has analyzed it with regard to taste. One of his generalizations is that mild states of stimulation produce pleasure (positive affect), intense states produce pain (negative affect), with the range of intensities involved dependent upon the specific sensory quality, such as sweet, sour, salty, or bitter (McClelland, 1951, pp. 467–470). Young (1959) has made a similar analysis based upon food preferences as related to concentration in solution.

It is postulated that certain conditions, which innately are affectively positive, form the constitutional base for the elaboration of sexual appetite:

1. Mild local stimulation of the genitals is innately pleasurable. This is a matter of common observation by parents. The random exploration of the body by the infant's hands soon leads to an undue amount of time spent stimulating the genital organs unless some preventive action is taken. The phenomenon is more easily noticed in the male child because of the greater accessibility to stimulation of the penis than the vulva and because of the observability of the phenomena of tumescence and erection consequent upon such stimulation. Adults of the Atimelang tribe of Alor, Indonesia, have capitalized upon the pleasurable character of such stimulation; they may deliberately stimulate the genitals of the child to soothe and quiet him (Kardiner, 1945, p. 131). This also occurs among the Hopi, Siriono, and Kazak (Ford & Beach, 1951). The pleasure accompanying genital stimulation (arousal) continues throughout life as an affective base for motivational development. Many stimuli, once neutral, may become associated with it, as a consequence functioning as cues leading to genital arousal, even in the absence of tactual stimulation itself.

2. The phenomenon of genital climax, together with its sequel, relaxation, is highly pleasurable. It is postulated that arousal short of climax is pleasurable, though not as gratifying as arousal leading to climax (see Kagan, 1955; Sheffield, Wulff, & Backer, 1951; Whalen, 1961, for experimental evidence in rats). Ford and Beach (1951), relying largely on data from Dickinson, maintain that masturbation in women perseverates among those who experience orgasm if no feelings of guilt or disgust ensue. While those who fail to achieve orgasm tend more often to quit, some persist in the practice for years without it.

In men, the pleasures of "girl watching," voyeurism, viewing "cheesecake" photography (Clark, 1952; Mussen & Scodel, 1955), and reading suggestive literature are well known, though participation rarely culminates in orgasm. Reference is made to Beach's (1956) tentative analysis of the sexual arousal mechanism (SAM) and the intromissive and ejaculatory mechanism (IEM). He suggests these as separate and sequential mechanisms in the male, and further suggests that the IEM has changed little in the course of evolution and is under poor voluntary control, but that the unique features of human sexuality are consequent upon the great lability of the SAM. Both the SAM and the IEM are here presumed to be highly charged with positive affect.

CHILDHOOD SEXUALITY

In some societies (e.g., Alorese, Pilaga of Argentina), masturbation develops early and is maintained unabated through the childhood years, sometimes being elaborated into mutual masturbation and other forms of sex play, both homosexual and heterosexual. These forms of erotic expression would seem to be learned through the association of the pleasurable affect with the tactual stimulation producing it. It must be noted, of course, that other sources of affect than the erotic are frequently involved in motivations toward sexual behavior. An examination of the data of Henry and Henry (1953) on Pilaga Indian children reveals the importance of several additional factors. Sex games are used to manifest hostility and dominance. Also, parental intercourse is openly viewed by the children from a very tender age. Here the child is excluded from a social experience of powerful affect, and it may well be that in both his doll play and interpersonal play, the Pilaga child is seeking to capture the warmth of inclusion in such an affect laden experience. While in the present article the primary focus is upon the erotic or lustful affect associated with sex, it must be remembered that this affective component is nearly always integrated with other sources of affect in the sexual behavior of humans of all ages.

In many societies, more or less active steps are taken to limit or prevent the masturbatory efforts of children, and to punish other presumed evidences of sexual curiosity or interest such as boy-girl exploration of genitals, public nudity, use of sex words, etc. The Victorian mores of nineteenth century Vienna are generally recognized today as the basis for Freud's postulated latency period, in which a dearth of conscious or overt sexual material occurs in (approximately) the sixth through the twelfth years. In his scheme the latency period follows closely upon the heels of the "phallic" stage in which the libido is presumed to be localized for the first time in the genitalia. This sequence is understandable, for it is precisely the behavioral evidences of the phallic stage noted above which evoke social sanctions of sufficient power to provoke inhibition or concealment of sexual expression. A wide range of sanctions may be employed, far beyond the veiled or explicit castration threats deemed to be so important by Freud. Verbal judgments, vague threats, corporal punishment, shaming, teasing, isolation, denial of privileges, distraction are some of the devices used (Sears, Maccoby, & Levin, 1957). In this way, expressions of sexual motivation may be largely prevented until the time of life deemed culturally appropriate.

From the present theoretical framework, the similarity of childhood sexuality to adult sexuality is basically a function of the comparative sophistication about sex based in affective experience, though without the potent affect associated with orgasm, a lesser childhood intensity of motivation would usually be expected. For the child to desire intercourse (e.g., with the opposite-sex parent), he would have to know about intercourse, would have to have positive affective expectations about it, and would have to perceive the person involved as a potential sex object. Among the Pilaga Indians, where such conditions seem to pertain, little difference in sexual motivation other than in intensity might be expected between children and adults. But great differences would be expected between the naive, carefully protected United States middle-class child and his parents. To impute to the child the sexual desires of the experienced adult is entirely unwarranted unless the child has the requisite background of affective learning.

Clinical studies by Levy (1937), Goldfarb (1944, 1945), Spitz and Wolf (1949), and others have suggested that extreme social deprivation in infancy produces severe and lasting disruption of affectional (and intellectual) processes, though others dispute this (Dennis & Najarian, 1957; Pinneau, 1955). The Harlow (1962) and the Harlow and Harlow (1962) ingenious studies of the development of affectional systems in monkeys give promise of identifying the critical variables (at least for monkeys!), though it is too early for results to be definitive. In the rhesus monkey, social deprivation in infancy appears to produce lasting inability to exhibit adequate mating behavior, but lack of peer relationships seems far more critical than the absence of the mother-child relationship. It should be noted that in Dennis and Najarian's study, the infants had active peer relationships after the first 6 months or so, though they were almost totally deprived of relationships with adults. Although these children seemed somewhat retarded in infancy, they appeared to be normal when tested at a later age (3–5 years).

In any event, there seems to be mounting evidence that extreme social deprivation in early childhood may produce pathological effects upon the development of "normal" heterosexuality, supporting the thesis that sexuality is in large part a derivative of experiential factors.

DEVELOPMENT OF SEXUAL APPETITE IN DATING AND COURTSHIP

In order to take a closer view of the development of sexual motivation, let us assume a "normal" but reasonably unsophisticated United States boy and girl who have entered our culturally prescribed period of heterosexual activities, the dating period. By this time, say at age 15, they each have gone through a long period of largely incidental learning about such "presexual" activities as holding hands, kissing, and embracing. Exposure to the mass media, "popular" songs, the romantic ventures of older siblings, teasing about girl friends or boy friends—indeed, myriads of influences have given rise to the development of a family of related expectations that engaging in behavior such as holding hands, walking arm in arm, kissing, and embracing with the "right" (culturally delimited) person leads to a positive affective experience. These activities are perceived as related to one another, and thus they constitute a family or system. In the absence of personal affective experience, these expectations are not very compelling in motivational strength; but they may, under the proper conditions, lead to the initiation of such behavior. These conditions would include the perception of the partner as an appropriate object for romantic interest, a social situation of privacy or intimacy where affectively negative consequences resulting from kidding, tattling, etc., are minimized, and a developing relationship between the couple such that the girl is perceived as providing encouragement to the attentions of the boy.

As the boy becomes more assured by the conditions mentioned to expect positive consequences from his attempt to hold the girl's hand, the motivations leading to such behavior will begin to outweigh those restraining such expression, and he will initiate the activity. Assuming he has appraised the situation correctly, his grasp of the girl's hand will be reciprocated by the girl. In our society the chances are excellent that this contact will produce a pleasurable emotional state, the occurrence of which confirms the previously held expectation, and also other members of the set of expectations.

Why does holding hands produce this emotional response? Obviously, it is not because of some innate connection between receptors in the hand and neural centers of affective arousal, erotic or otherwise, since equivalent stimuli from other persons do not produce this response at all. Rather, this reaction is produced because of the learned symbolic significance of holding hands with the appropriate person. And there need not be any sexual connotation attached. This special intimacy may gratify a need for psychological closeness with someone; or it may fulfill needs for status building by meriting such a relationship with a person favorably placed in the status system of adolescent culture. Simply put, it is satisfying to be able to hold hands with someone perceived as attractive (in the modern slang, "real cool"), and doubly satisfying if this person is generally so regarded by one's age-mate reference group. The importance of esteem motives in dating has been described for an earlier generation by Waller (1937), and more recently has been graphically portrayed by Margaret Mead (1949, pp. 281–295). Other motives may be involved, of course, but the point is that initial participation in hand holding and further romantic activities generally originates in nonsexual motives, and further, that the positive affect arising from the activity serves to strengthen motivation for it.

As the dating period continues, the tendency for the couple to engage in kissing is fostered by the increased motivational strength generalized from hand holding, also by the growing intimacy of the relationship which alters the distribution of positive and negative expectations, e.g., the partner permitting or encouraging such a thing. Assuming that the kiss is attempted and is mutually acceptable, a positive affective state is likely to result, with consequences comparable to those discussed above for hand holding. If the girl's lipstick tastes good, there may be another reward, which Young (1952) would call the "palatability effect"!

But an additional development is going on at the same period of time. The couple is getting more educated about things sexual through a more sophisticated exposure to the "facts of life" via reading, age-mate conversations, and a host of other sources. So a new set of expectations is established about "petting," sexual intercourse, and the like, with affect more or less vividly imagined. This new set may have only a tenuous relationship with the former set initially, a bond which will become firmer as the association value between the sets increases.

At this point, application needs to be made of the principle of habituation, which states that the gratification which comes from an initial experience occurs with lessened intensity upon subsequent occasions. The dulling of the thrill with repetition may lead the person or couple to try out new activities in order to regain the prior level of gratification. In the case of the hypothetical couple this habituation effect heightens the likelihood that they will resort to progressively more intense love-making, e.g., embracing, "necking," petting, as they continue to engage in amatory behavior.

The confirmation of their affective expectations, the increase of their romantic appetite with gratification, and the dulling of the thrill of any given activity with habituation all combine to move the couple along the path of more affect laden experiences.

Now at some point in this sequence there will occur a true sexual arousal (lust). The specifically sexual connotations of certain acts, e.g., kissing the ear, fondling the breasts, indeed, a host of verbal or gestural cues of affection or assent, can set off this reaction. Once this erotic arousal has occurred, bringing powerful excitement in its wake, the desire to repeat and continue the experience is greatly enhanced; thus the sexual appetite is increased. Another important event also takes place, however. A new set of associations or meanings is formed. The activities which led up to the initial erotic arousal now have a tendency to serve as cues leading to the arousal of sexual desire. Furthermore, as erotic experiences are repeated (*a*) the greater the association value of these cues to sexuality, and (*b*) the wider the range of cues to sexuality. Thus, kissing, embracing, exposure of body parts ordinarily covered, and a host of other acts may serve as sexual stimuli to the person with a history of arousal, while to the naive person they have no such significance. A 9-year-old, for example, may find intimate scenes in the movies boring while they may be highly arousing to his experienced elders.

Summarizing, the positive affect accompanying sexual arousal increases the appetite for such experiences; and the occurrence of such experiences produces an association of the preceding stimuli to the sexual affect, thereby making these stimuli cues leading to the arousal of sexual desire. These developments, together with the habituation phenomenon, tend to carry the participants toward the point where the couple may become so excited that it is virtually impossible for them to stop short of climax. In Beach's terms, the IEM may become sufficiently aroused for the male that it carries itself to completion with virtually no voluntary control. This may occur without actual intromission (petting to climax), though intercourse will tend to follow upon subsequent occasions due to the operation of the factors just mentioned.

The careful study of Ehrmann (1959) provides much data pertinent to the behavior discussed above, and also the problems of approach-avoidance and control of sexuality discussed below. His data are not sufficiently precise to confirm or deny aspects of the theory. His results are generally consistent with appetitive theory, though admittedly they are subject to alternative interpretations. To test precise details of any theory of sexual behavior, it will be necessary to go beyond his study and those of Kinsey et al. (1948, 1953) to two kinds of research: that which is more experimental in character, providing necessary control of variables, and that which studies individual cases more intensively on a longitudinal basis.

PROBLEMS OF APPROACH AND AVOIDANCE

To this point, only the positive affect has been considered. However, there are several sources of negative affective expectations surrounding sexual behavior which need to be considered to give an adequate treatment of sex motivation.

One important source is the guilt arising out of moral compunctions against various expressions of sexuality. These arise in varying degree among persons in the great bulk of societies throughout the globe. A second source is the fear of discovery, of bringing disgrace to one's family and ruination to one's reputation. For the unmarried, a third source, the fear of pregnancy, is compounded from the guilt and fear already mentioned, together with the loss of freedom accompanying a forced marriage. For both the married and unmarried other anxieties attendant to pregnancy may arise: threats to health, the stress of psychological and economic adjustments to the baby. A fourth source of concern is the fear of being "used," with the loss of interest by the partner once he has gotten what he wanted. A fifth source common to women is the disgust and revulsion toward sex produced by the doctrine that sex is filthy and degrading (French Institute of Public Opinion, 1961, pp. 159–163).

The above list of sources of negative affect is not meant to be exhaustive, but to lend credence to the proposal that one or both partners in a sexual relationship may have potent ambivalent feelings toward it.

A Lewinian analysis of driving and restraining forces (Lewin, 1947) is useful in describing the approach-avoidance dynamics involved in sexual conflict and the management of sex appetite.

In American society, the burden of the responsibility for the control of sexuality rests upon the female. In the dating and courtship period, a girl may decide that she will allow a boy to go only so far, and then she will "draw the line." Thus, behavior will occur at a given level because of the balance of driving forces with restraining forces. This is an unstable equilibrium, however, and the device therefore ineffective if love-making is habitual up to this point, since the increased appetite may overbalance the restraints, causing a shift in the line. The resulting guilt, anxiety, and cognitive dissonance are then assuaged by rationalization and other available defenses.

One prevalent manifestation of ambivalence about sex is the divorcement of sexual qualities and feelings from conceptions of the "good" mate. Thus a young man may coax a girl to grant sexual favors, then lose interest in her as a prospective wife when she does, since he cannot conceive of marrying a girl who is promiscuous! Because "love is wonderful but sex is nasty" the girl, too, is likely to perceive the boy who wants to make love as one who is untrustworthy, one who is perhaps not interested in her but only in what selfish satisfaction he can obtain. Consequently, while the couple is verbally reassuring one another of their affection while they make love, they actually experience mounting doubts about the genuineness of their devotion as the frequency and intensity of the love-making increase. Both before and after marriage the enjoyment of sex may be colored by a loss of respect for the partner as one who is no longer pure and sweet. This occurs because the good person, shorn of sexual quality, is overidealized. When erotic characteristics are discovered in that person, disillusionment results.

The very personal gratification which accompanies sexual activity lies at the heart of the enigma of love and sex, since it raises the eternal question: Is the other person making love to obtain selfish satisfaction or out of a more inclusive feeling for me and an altruistic concern for my happiness? The dissociation of love and sex just discussed is one attempt to resolve this problem. Another is the doctrine that love and sex are virtually synonymous. While this theory typifies much of the wishful content of the mass media today, physical (sexual) attraction is probably not convincing to the lovers themselves as a criterion of love.

Reiss (1960), in his analysis of American premarital sex standards, suggests that there is a generally recognized distinction between sex for sex' sake (body centered coitus) and sex for love's sake (person centered coitus). Most American women reserve intercourse for lovers, as an affectionate expression of love. The quandary which exists is not merely whether the woman "really" loves the man, but whether the man really loves her. If the man is sexually opportunistic, aggressive, or demanding—and a wealth of data (e.g., Ehrmann, 1959; Kinsey, Pomeroy, & Martin, 1948; Reiss, 1960) suggests that the majority of American men are—then the woman is left to seek for nonsexual assurances that he really does love her. The more insistent he is about sexual relations, the more unsure of his love she becomes. If they are not married, the uncertainty is magnified, because of the lack of marital responsibilities assumed by the lover.

A specialized consequence of ambivalence about sex is the development of fetishism or other bizarre aberrations. The case described by Grant (1953) may be chosen as an illustration of this, though his data are insufficient to make any analysis certain. In such cases, a dissociation occurs between one's erotic desires and more legitimate, upright impulses, and a corresponding split develops between sexual and lofty qualities perceived in others. Thus, in Grant's case, the patient in earlier years could have sexual escapades with a woman of easy virtue and a platonic relationship with a good woman toward whom no erotic interest could be tolerated; and "never the twain shall meet." The patient's resolution of his problem by developing a fetish about shoes, legs, and leg movements may be interpreted as an attempt to control deep-lying feelings of guilt about licentious relationships with women. In view of this, his final solution, which consisted of viewing art films emphasizing leg displays, seems rather clever because no one else is involved, guilt is thereby minimized, and his own risk of exposure is virtually eliminated.

Cases of sexual assault, exhibitionism, and the like may result from the eruption of sexual desire through the barriers imposed by morality and reality testing. The development of such a breakthrough would be explained in the following fashion. The person engages in a period of lustful fantasy, in many cases stimulated by romantic or pornographic material. The excitement which accompanies this activity is sufficiently rewarding to increase the appetite. If such stimulation occurs over a protracted period, a greatly aggravated appetite may ensue. If socially acceptable channels of expression are denied the person, he may more or less planfully enter a social setting of erotic outlet where the chances for detection and capture are reduced. Thus the avoidance forces deterring him from such an act, reduced by lowered expectations of negative consequences, are overwhelmed by the mounting approach tendencies, which increase as the lascivious longing continues.

VALUATIVE APPLICATIONS OF THE THEORY

A brief attempt will now be made to apply appetitional theory to problems of the personal and social management of sexuality, as follows: premarital relations, masturbation, control of sexual appetite, and sexual compatibility in marriage.

Premarital Relations

Some observers have argued that American premarital sex mores should be made more lenient. Their argument is based upon evidence that there is widespread violation of these standards. Such a position assumes that the standards are more flexible than the behavior; therefore, the standards should be altered to conform to the realities of conduct. If one assumes that libido is a function of biological sex maturation, therefore subject to limited control, then since American culture tends to delay marriage for a number of years beyond sexual maturity, it follows that the violation of standards of premarital chastity is virtually inevitable.

However, if one assumes that biological maturation is only a minor factor in the development of sexual desire, and that sex appetite is subject to a tremendous degree of control, then other conclusions are permitted. Furthermore, if there are negative personal-social consequences of premarital indulgence which outweigh the positive consequences, as suggested by the Judeo-Christian moral code, then one might conclude that the behavior should be altered rather than the standards.

If such a conclusion be entertained, the question then follows: On what does the behavior depend, and how may it be altered? Based on the above analysis of the development of sexual appetite, several factors can be identified as contributive to the growth of libido: (*a*) Early dating where presexual behavior is permitted and even encouraged. (*b*) "Going steady" early in the courtship years, which tends to increase the likelihood of the next factor. (*c*) A dating pattern which is characterized by chronic love-making. (*d*) The bombardment of adolescents with sexually stimulating materials of many kinds, such as suggestive or obscene literature, sexy motion pictures, etc. It might be noted that the drive-in theater is called a "passion pit" not because of climatological factors which generate gonadal secretions, but because the lustful pictures frequently shown stimulate and legitimize sexual expression, and the darkened privacy of the automobile lowers other barriers to such behavior. (*e*) The lowering of moral standards which serve to restrain immediate gratification of impulse. Emphasis upon and idealization of glamor, physical attractiveness, and fun tend to produce a decreased concern with other values which depend upon long-term action. The absence of moral training in a setting of sex sophistication leaves a crucial void in restraining forces.

Masturbation

The use of self-stimulation as a source of erotic pleasure is reported to be very widespread among American adolescent males, and also to be common among females (Kinsey, Pomeroy, & Martin, 1948; Kinsey, Pomeroy, Martin, & Gebhard, 1953). The recognition of this has led to a curious reversal of a former position. Whereas previous generations were warned not to masturbate because they might thereby become insane or abnormal, many argue today that those who fail to indulge in this practice must somehow be unbalanced or abnormal. The limitations of such a statistical conception of normality are too well known to require comment.

The present generation would seem to be "victims" of the misconception that periodic sexual outlet is biologically required. This notion is very handy, and is readily linked to the feeling that one might as well enjoy it, controlling evocation of the impulse so as to maximize pleasure.

Needless to say, where moral standards exist which forbid such gratification, approach-avoidance conflict exists. As in the case of premarital intercourse, conflict can be resolved by changing standards. However, it may be preferable to alter the behavior to fit the standards.

Control of Sexual Appetite

A device sometimes suggested to control sexuality is to reduce tension by displacement or sublimation, i.e., to relieve sexual drive by reading, dancing, painting, attending the theater, etc. According to appetitional theory, if the substitute activity is sexually stimulating, then the appetite will tend to increase, and the problem of control is worsened. If the activity is nonsexual in character, sexual impulses will not be relieved nor aroused. If a person finds enough such activities which are satisfying because of nonsexual reasons, his need for sexual gratification will be lessened, not because of sublimation of libido, but simply because the role of sexual gratification in the total motivational economy of the personality will be relatively minor. In contrast to Freud's view that much human activity gratifies a sublimated libido, it seems probable that sex is often indulged in as a substitute for a reasonably satisfying life in other areas.

Appetitional theory suggests a set of recommendations for the control of lust during the premarital period.

1. Participation in activities which fill one's life with abundant gratifications, so that sexual intimacies are not indulged in as an escape from boredom or emotional starvation. Athletic, intellectual, altruistic, occupational, avocational, esthetic, or social pursuits may serve this function.

2. Comprehension of the nature of sex motivation and the factors which govern it, to help the person to foresee the consequences of alternative courses of action, thus enabling him to make a knowledgeable decision in terms of his own motives and values.

3. Prevent the appetite from growing by abstinence from arousing activities until the time acceptable for sexual expression. This implies a dating-courtship pattern which is centered upon a wide variety of nonsexual activities which additionally might contribute to a more adequate prediction of marital compatibility. (See Leuba, 1954, for a related set of recommendations.)

4. Exposure to entertainment, reading material, companionship, etc., which have nonsexual connotations of intrinsic interest to the person. This will tend to prevent the recurrent arousal of sexuality.

5. Emphasis upon the long-term rewards of continence, and idealization of the ability to delay immediate gratification of impulse.

The implication should not be drawn from the preceding suggestions that an overprotective, authoritarian control of behavior is favored. The emphasis recommended is on self-selection of situations and activities. Of course, in a society as exploitative as America, some degree of social control of blatant obscenity is probably required.

Sexual Compatibility in Marriage

The problem of the role of sexual compatibility in marriage is another area of application of the theory. It is sometimes maintained that sexual compatibility is of central importance to a successful marriage. Based upon this argument is the notion that a couple needs to have intercourse before marriage to find out if they are compatible.

Appetitional theory supports a contrasting view. Rather than conceiving of a person's sexual drive as a relatively fixed quantity which can only be compared with that of another individual, the present theory maintains that the appetite is a variable quantity which might be adjusted over a period of time to match that of another. Thus a couple whose initial appetites are at variance may work toward a mutual adjustment of these until greater harmony exists. Their willingness to do this probably depends upon their nonsexual love for one another.

SUMMARY

1. An appetitional theory of human sexual motivation is proposed as a needed alternative to theories based

upon sex as a biological drive or instinct in view of the mounting evidence against the adequacy of physiological interpretations.

2. The theory is based upon a general motivational orientation which maintains (*a*) the affective consequences of behavior are central in the development of motives, through the mediation of associations between cues and affect. These cues, perceptual or cognitive, tend to arouse expectations that a given pattern of action will produce a certain affective change. For an expectation to be motivational in character, it must involve not only affective change, but also a course of behavior, symbolic or overt, which is instrumental to attaining the desired affect; (*b*) motives are strengthened in proportion to the degree of affective change experienced; (*c*) the likelihood of arousal of a motive is a function of the likelihood of occurrence of cues to the affective change and the associative strength of the cues to the affect; (*d*) all members of a family or set of expectations are reinforced by the confirmation of one member of the set along a gradient of generalization; (*e*) a habituation process occurs in many affective areas wherein the repetition of a given activity produces a diminished affective change in comparison with the initial experience; (*f*) approach motives are based upon positive affective expectations, avoidance motives upon negative affective expectations; however, most motivated behavior is ambivalent, based upon a mixture of positive and negative components; (*g*) of a number of expectations aroused at any time, the one acted upon depends on the relative positive affective change anticipated over a period of time, and the subjective probability of attainment of the alternatives.

3. The affective foundations for genital sex motivation (lust) are (*a*) the pleasurable excitement of arousal, and (*b*) the intense pleasure accompanying orgasm and the relaxation which follows. Either of these sources of affect is adequate to strengthen sexual appetite. It is recognized that there are many nonlustful motives which contribute to participation in sexual behavior.

4. Childhood sexuality is based primarily upon the first affective foundation, and becomes more potent with indulgence; however, in many societies the precautions taken to prevent arousal together with various threats of negative consequences are usually adequate to keep such behavior subdued to a rather inconsequential level.

5. The burgeoning of sexual motivation usually occurs during the culturally prescribed period of courtship. Middle-class American society is used as an illustration. The person or couple engages in amatory adventures of increasing intensity due to the effect of reinforcement, generalization of reinforcement, and habituation. At some point in this sequence true genital arousal (lust) occurs, its affective power greatly heightening the motivational level. Chronic participation in arousing activity augments the appetite, which overwhelms even strong resistances composed of moral compunctions and various fears, eventually leading to intercourse.

6. The theory is applied to problems in the management of sexual appetite in premarital courtship and masturbation and to the role of sexual compatibility in marriage.

REFERENCES

ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.,* 1957, **64**, 359–372.

BALL, JOSEPHINE, & HARTMAN, C. J. Sexual excitability as related to the menstrual cycle in the monkey. *Amer. J. Obstet. Gynec.,* 1935, **29**, 117–119.

Beach, F. A. Evolutionary changes in the physiological control of mating behavior in mammals. *Psychol. Rev.*, 1947, **54**, 297–315.

Beach, F. A. Instinctive behavior: Reproductive activities. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 387–434.

Beach, F. A. Characteristics of masculine "sex drive." In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1956.* Lincoln: Univer. Nebraska Press, 1956. Pp. 1–32.

Benedek, Therese, & Rubenstein, B. B. The correlations between ovarian activity and psychodynamic processes. *Psychosom. Med.*, 1939, **1**, 245–270; 461–485.

Benedek, Therese, & Rubenstein, B. B. The sexual cycle in women. *Psychosom. Med. Monogr.*, 1942, **3**, 1–306.

Bruner, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald Press, 1951. Pp. 121–147.

Carmichael, H. T., Noonan, W. J., & Kenyon, A. T. The effects of testosterone propionate in impotence. *Amer. J. Psychiat.*, 1941, **97**, 919–943.

Cartwright, D. Decision-time in relation to the differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442.

Cartwright, D., & Festinger, L. A quantitative theory of decision. *Psychol. Rev.*, 1943, **50**, 595–621.

Clark, R. A. The projective measurement of experimentally induced sexual motivation. *J. exp. Psychol.*, 1952, **44**, 391–399.

Cleckley, H. *The caricature of love.* New York: Ronald Press, 1957.

Cory, D. W. Homosexuality. In A. Ellis & A. Abarbanel (Eds.), *The encyclopedia of sexual behavior.* New York: Hawthorn, 1961. Pp. 485–493.

Creevy, C. D., & Rea, C. E. The treatment of impotence by male sex hormone. *Endocrinology*, 1940, **27**, 392–394.

Davis, Katharine B. *Factors in the sex life of 2200 women.* New York: Harper, 1929.

Dember, W. N., & Earl, R. W. Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychol. Rev.*, 1957, **64**, 91–96.

Dennis, W., & Najarian, Pergrouhi. Infant development under environmental handicap. *Psychol. Monogr.*, 1957, **71**(7, Whole No. 436).

Dickinson, R. L., & Beam, L. *A thousand marriages.* Baltimore: Williams & Wilkins, 1931.

Ehrmann, W. *Premarital dating behavior.* New York: Holt, 1959.

Ellis, A. The sexual psychology of human hermaphrodites. *Psychosom. Med.*, 1945, **7**, 108–125.

Ellis, H. *Studies in the psychology of sex.* Vol. 1. New York: Random House, 1936. Pp. 97–105.

Filler, W., & Drezner, N. The results of surgical castration in women under forty. *Amer. J. Obst. Gynec.*, 1944, **47**, 122–124.

Finesinger, J. E., Meigs, J. V., & Sulkowitch, H. W. Clinical, psychiatric, and psychoanalytic study of a case of male pseudohermaphroditism. *Amer. J. Obst. Gynec.*, 1942, **44**, 310–317.

Ford, C. S., & Beach, F. A. *Patterns of sexual behavior.* New York: Harper, 1951.

French Institute of Public Opinion. *Patterns of sex and love: A study of the French woman and her morals.* New York: Crown, 1961.

Gerall, A. An attempt to induce precocious sexual behavior in male guinea pigs by injections of testosterone propionate. *Endocrinology*, 1958, **63**, 280–284.

Goldfarb, W. Infant-rearing as a factor in foster home replacement. *Amer. J. Orthopsychiat.*, 1944, **14**, 162–166.

Goldfarb, W. Effects of psychological deprivation in infancy and subsequent stimulation. *Amer. J. Psychiat.*, 1945, **102**, 18–33.

Grant, V. W. A case study of fetishism. *J. abnorm. soc. Psychol.*, 1953, **48**, 142–149.

Hamilton, G. V. *A research in marriage.* New York: Boni, 1929.

Hamilton, J. B. Demonstrated ability of penile erection in castrate men with markedly low titers of urinary androgens. *Proc. Soc. Exp. Biol. Med.*, 1943, **54**, 309–312.

Harlow, H. F. The heterosexual affectional system in monkeys. *Amer. Psychologist*, 1962, **17**, 1–9.

Harlow, H. F., & Harlow, Margaret K. Social deprivation in monkeys. *Scient. American*, 1962, **207**, 136–146.

Henry, J., & Henry, Zunia. Doll play of Pilaga Indian children. In C. Kluckhohn, H. Murray, & D. M. Schneider (Eds.), *Personality in nature, society, and culture.* (2nd ed.) New York: Knopf, 1953. Pp. 292–307.

Johnson, D. M. The phantom anesthetist of Mattoon: A field study of mass hysteria. *J. abnorm. soc. Psychol.*, 1945, **40**, 175–186.

Kagan, J. Differential reward value of incomplete and complete sexual behavior. *J. comp. physiol. Psychol.*, 1955, **48**, 59–64.

KARDINER, A. *The psychological frontiers of society.* New York: Columbia University Press, 1945.

KINSEY, A. C. Criteria for a hormonal explanation of the homosexual. *J. clin. Endocrinol.*, 1941, **1**, 424–428.

KINSEY, A. C., POMEROY, W. B., & MARTIN, C. E. *Sexual behavior in the human male.* Philadelphia: Saunders, 1948.

KINSEY, A. C., POMEROY, W. B., MARTIN, C. E., & GEBHARD, P. H. *Sexual behavior in the human female.* Philadelphia: Saunders, 1953.

LEUBA, C. *The sexual nature of man and its management.* New York: Doubleday, 1954.

LEVY, D. M. Primary affect hunger. *Amer. J. Psychiat.*, 1937, **94**, 643–652.

LEWIN, K. Behavior and development as a function of the total situation. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946. Pp. 791–844.

LEWIN, K. Frontiers in group dynamics: Concept, method, and reality in social science; social equilibria and social change. *Hum. Relat.*, 1947, **1**, 5–42.

LEWIN, K. *Resolving social conflicts.* New York: Harper, 1948.

MCCLELLAND, D. C. *Personality.* New York: Sloane, 1951.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MEAD, MARGARET. *Male and female.* New York: Morrow, 1949.

MILES, CATHERINE C. Psychological study of a young male pseudohermaphrodite reared as a female. In Q. McNemar & Maud Merrill (Eds.), *Studies in personality.* New York: McGraw-Hill, 1942. Pp. 209–227.

MILLER, N. E. Comments on theoretical models illustrated by the development of a theory of conflict behavior. *J. Pers.*, 1951, **20**, 82–100.

MONEY, J. Hermaphroditism. In A. Ellis & A. Abarbanel (Eds.), *The encyclopedia of sexual behavior.* New York: Hawthorn, 1961. Pp. 472–484.

MONEY, J., HAMPSON, JOAN G., & HAMPSON, J. L. Hermaphroditism: Recommendations concerning assignment of sex, change of sex, and psychologic management. *Johns Hopkins Hosp. Bull.*, 1955, **97**, 284–300. (a)

MONEY, J., HAMPSON, JOAN G., & HAMPSON, J. L. An examination of some basic sexual concepts: The evidence of human hermaphroditism. *Johns Hopkins Hosp. Bull.*, 1955, **97**, 300–319. (b)

MONEY, J., HAMPSON, JOAN G., & HAMPSON, J. L. Imprinting and the establishment of gender role. *AMA Arch. Neurol. Psychiat.*, 1957, **77**, 333–336.

MUSSEN, P. H., & SCODEL, A. The effects of sexual stimulation under varying conditions on TAT sexual responsiveness. *J. consult. Psychol.*, 1955, **19**, 90.

NISSEN, H. W. Instinct as seen by a psychologist. *Psychol. Rev.*, 1953, **60**, 291–294.

PINNEAU, S. R. The infantile disorders of hospitalism and anaclitic depression. *Psychol. Bull.*, 1955, **52**, 429–452.

REISS, I. L. *Premarital sexual standards in America.* Glencoe, Ill.: Free Press, 1960.

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

SEWARD, GEORGENE H. *Sex and the social order.* New York: McGraw-Hill, 1946.

SHEFFIELD, F. D., WULFF, J. J., & BACKER, R. Reward value of copulation without sex drive reduction. *J. comp. physiol. Psychol.*, 1951, **44**, 3–8.

SPITZ, R. A. & WOLF, KATHERINE M. Autoerotism. In, *Psychoanalytic study of the child.* Vol. 3/4. New York: International Universities Press, 1949. Pp. 85–120.

TAUBER, E. S. Effects of castration upon the sexuality of the adult male: A review of relevant literature. *Psychosom. Med.*, 1940, **2**, 74–87.

TERMAN, L. M. *Psychological factors in marital happiness.* New York: McGraw-Hill, 1938.

TINKLEPAUGH, O. L. The nature of periods of sex desire in women and their relation to ovulation. *Amer. J. Obst. Gynec.*, 1933, **26**, 335–345.

WALLER, W. The rating and dating complex. *Amer. sociol. Rev.*, 1937, **2**, 727–734.

WHALEN, R. E. Effects of mounting without intromission and intromission without ejaculation on sexual behavior and maze learning. *J. comp. physiol. Psychol.*, 1961, **54**, 409–415.

YOUNG, P. T. The role of hedonic processes in the organization of behavior. *Psychol. Rev.*, 1952, **59**, 249–262.

YOUNG, P. T. The role of affective processes in learning and motivation. *Psychol. Rev.*, 1959, **66**, 104–125.

(Received December 27, 1961)

Psychological Review
1964, Vol. 71, No. 1, 19–26

NEUROPHYSIOLOGICAL CONTRAST PHENOMENA AND FIGURAL AFTEREFFECTS[1]

J. A. DEUTSCH
Stanford University

It is proposed that figural aftereffects are produced by the same neurophysical mechanism as are simultaneous and successive contrast. Observations on lateral inhibition in the eye of *Limulus* show that after an ommatidium has ceased to be more highly illuminated than its neighbors, it undergoes a postexcitatory depression and its neighbors a postinhibitory rebound. If in addition we assume that the apparent distance between 2 points is measured in the nervous system by summing the excitability of the elements between them, then distortions of phenomenal space will occur around previously excited points. These are seen to correspond to those observed in figural aftereffects, e.g., in showing a distance paradox. Various consequences are discussed.

Writing of their elegant theory of inhibitory interaction in the *Limulus* eye, Hartline, Ratliff, and Miller (1961) state,

> The value of such a theoretical formulation lies in the compactness with which a number of experimental observations are described, and the insight it gives into the understanding of relationships that, while inherently simple, sometimes yield phenonmena of considerable complexity. Applied with caution, it may be useful as a basis for hypothetical explanations of such phenomena as brightness contrast and sensitivity to moving patterns in human vision.

In this paper, it is hoped to show that figural aftereffects in the human visual system stem from neural properties qualitatively similar to those found in the eye of *Limulus*. Systems showing inhibitory interactions like those found in the *Limulus* eye have been described by Kuffler (1953), Hubel and Wiesel (1959), and Baumgartner (1961) in the visual system of the cat, though here such systems show features of added complexity. As the present hypothesis assumes only the inhibitory interactions found in the *Limulus* eye in explaining a phenomenon in human vision, it may well turn out to be oversimplified, though a theory any more elaborate would hardly be justified in our relatively primitive state of knowledge.

THE HYPOTHESIS

The basic assumptions of the present hypothesis are mainly those which already have been observed in many sensory systems (see review by Brooks, 1958). It is assumed that an excited element inhibits other less excited elements, the degree of such inhibition falling off with distance. Furthermore, it is assumed that when the excited element ceases to be excited, its excitability will transiently diminish below its previous baseline, and as a result, the elements, which were previously inhibited by its activity, will now undergo a corresponding transient increase in excitability. Such a relation has been shown in the eye of *Limulus* and is illustrated in Figure 1 (Ratliff, 1961).

---

[1] The work was supported by United States Public Health Service Grant M-4563 and National Science Foundation Grant G21376.

My thanks are due to A. H. Hastorf for valuable discussions.

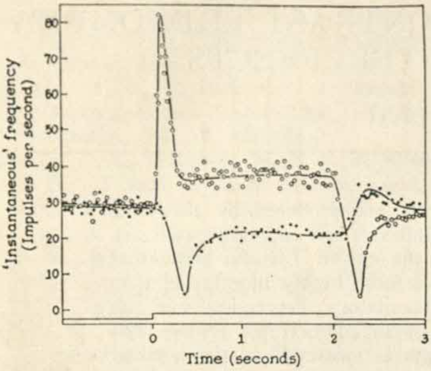


FIG. 1. Two adjacent ommatidia in the eye of *Limulus* and their inhibitory interaction. (One [black filled circles] was illuminated steadily throughout the entire period on the graph. The other ommatidium [open circles] was illuminated steadily until the point marked *O* on the graph. A sudden increment of illumination, remaining steady for two seconds, then occurred. At the end of the two seconds, its illumination was abruptly decreased to its original level. Of special interest here is the transient increase in frequency over the original steady state of the first ommatidium [black filled circles] which lags the transient depression in activity of the second ommatidium when its increment of illumination is abruptly turned off.)

These assumptions were in their general form proposed almost a century ago to explain simultaneous and successive contrast in human vision (Mach, 1865), and in more quantitative form yielded interesting derivations of such phenomena as "Mach patterns" (Ratliff & Hartline, 1959), illusions "seen" by *Limulus,* as well as perceived by man. An attempt has also been made to explain certain streaming phenomena on a similar basis (Taylor, 1958). In this paper, an explanation is proposed of figural aftereffects which also uses the above assumptions. Figural aftereffects first were described by Gibson (1933) and further investigated by Köhler and Wallach (1944). The whole subject has been well reviewed by McEwen (1958). In this illusion, contours that are placed close in the visual field to the locus of a previously exposed contour tend to be phenomenally shifted away from this locus. The extent of this shift is largest not when the two contours—the one that was originally fixated (the inspection figure), and the one which subsequently was substituted (the test figure)—are closest together; instead, the maximum effect of displacement is secured when the two contours occupy loci at a small intermediate distance from each other. (This is called the "distance paradox.") Not only does continuous fixation of the inspection contour give rise to a displacement of the second or test contour, but paling and receding from the observer are also reported. How far such a phenomenal shift away from the observer is due to paling is difficult to establish.

In order to explain such apparent changes in visual distance and relate them to the above assumptions concerning contrast, we must make one further hypothesis. We shall assume that distances between points are measured in the nervous system by taking some kind of index of excitability of the points lying between them. Lowering the excitability of the intervening points will lead to an overestimation of the distance, and increasing the excitability of such points will lead to an underestimate. This concept may be clarified by use of an everyday analogy. It is convenient when traveling, say, in an airplane, to estimate distance by the time that it takes one to travel between two points. However, local conditions such as headwinds or tailwinds (to correspond to states of lessened and heightened excitability) occurring over portions of our journey, may lead to overall overestimates or underestimates of the total distance. As the precise mechanism by which the sum of excitabilities of the points between two loci is measured is immaterial to our prediction of figural aftereffects, we need not embark on a discussion of such possible mechanisms here, but postpone it to a latter section.

Taking then our further assumption and combining it with those regarding neurophysiological contrast, if we place two points in a region close to a previously excited locus, some interesting consequences emerge. If the two points are placed so that the whole of the space between them is an area previously inhibited by the formerly excited locus, then they will seem closer together. The previously inhibited area will now be temporarily more excitable and so distances measured across it will be underestimated. On the other hand, if the two points are placed one each side of the previously excited locus, close to its edge, then they will seem to be farther apart. The previously excited area will now be temporarily less excitable and so distances measured across it will be overestimated. How far each side of the previously excited locus we must go before underestimation due to enhanced excitability counteracts the overestimation caused by lessened activity, is difficult to predict, because of the statistical distribution of the initially excited locus due to small eye movements during fixation (Ratliff & Riggs, 1950). It can be seen by reference to Figure 2 how the hypothesis predicts further properties of figural aftereffects, such as the distance paradox. Here the dotted line *X* represents the deviation from the normal excitability which would be expected from the assumptions about neural lateral inhibition and its rebound. Depression of excitability occurs maximally at the locus which had been excited by the point of a contour of the inspection figure.

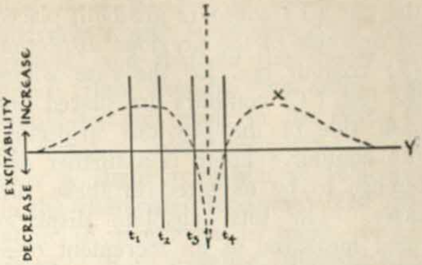


FIG. 2. Alterations in excitability of receptive layer due to the previous presence of inspection point at I. (*Y* indicates normal level of excitability; *X* the deviation from it. Alterations in apparent distance are computed by the area under the curve *X* lying between two test points, such as T_1, T_2 T_3, and T_4.)

If contours are placed at T_1, T_2, and T_3, since excitability in the regions between them is increased, apparent distance between them is decreased.

Further, the apparent shift of T_1 and T_2 toward each other will be larger than the shift of T_2 toward T_3, as the total amount of increase of excitability is greater between T_1 and T_2 than between T_2 and T_3. Hence, the largest shift does not occur nearest to the inspection (I) contour, but some distance away from it. In this way, the theory predicts the distance paradox first reported by Köhler and Wallach (1944). If two contours are placed each at T_3 and T_4, one each side of the previous inspection (I) contour, they will appear further apart since excitability in the region between them is decreased. The total absolute displacement of these contours will be larger than if they had been placed closer to the inspection (I) contour, thus yielding a distance paradox for this case also. However, it should be clear that the point at which displacement is maximal differs on this theory for the case where the test (T) contours are both placed on one side of the previous inspection (I) contour and for the case where the test (T) contours are placed one each side of the previous inspection (I) contour. There is a further difference to be expected in these two cases. The total absolute displacement measured as an increment over real distance for the case where the test contours are placed one each side of the inspection contour is greater than if they were placed closer to the locus of the previous inspection figure. However, if they were placed closer, their proportional displacement, relative to their real distance from each other, would be larger. This is not true of the two test contours, both placed on one side of the inspections contour. Here the proportionate shift (proportion by which space appears to shrink) will increase with distance from the inspection contour up to a certain point. Furthermore, both the absolute and percentage amount of shift ought to decline as we move one of the points of a pair on each side of the inspection figure further away from the other point. In fact, given an approximately equal influence of the increase in excitability and its decrease, we might expect any shift due to figural aftereffects to disappear completely as we moved both inspection points away from the locus of the test figure. On the other hand, as long as one of the points of a pair both on the same side of an inspection figure is fixed in an area of heightened excitability, the absolute, though not percentage, apparent decrement in distance from its partner should not decrease. (However, both the percentage and absolute decrement in distance may increase in this case if the increase in distance is taking in territory with a heightened excitability.)

As Fox's (1951) and Heinemann and Marill's work (1954) suggests, figural aftereffects can occur over large distances. On the present theory, we should expect to find a diminution of distance between two test points both on the same side of a test contour as long as one of them was some way within the territory previously covered by lateral inhibition from the test contour, but not if it was outside it. Some measurements of this type already have been undertaken (e.g., Bevan, 1951; Craik & Zangwill, 1939; Fry & Bartley, 1935). None has directly compared the distance over which figural aftereffects can occur with the extent of lateral inhibition or contrast within the same experiment. However, the distance over which thresholds are affected by the presence of previously present contours was found by Bevan (1951) to be 6 degrees and this is probably not the maximum figure. Therefore, we could on the present theory expect detectable figural aftereffects over considerable angular distances.

Hebb (1949) has pointed out that two test lines parallel to each other and at right angles to the inspection contour are seen closer to each other, whereas on the Köhler-Wallach (1944) theory, they should appear further apart.

On Osgood and Heyer's (1951) theory (discussed in greater detail elsewhere, Deutsch, 1956) the distance between them should suffer no distortion. The present theory would indeed predict that the two lines should be seen closer together, because they lie in a zone of enhanced excitability.

Köhler (1951) states that figural aftereffects seem to grow during the first few moments of the test period. That this should be the case is evident from the plot of the onset of increase of activity in a *Limulus* receptor unit after an adjacent unit has ceased to be stimulated. There is a lag of over .25 second before increase of activity reaches a maximum. Krauskopf (1954), who attempted to investigate the statement of Köhler's quoted above, and with negative results, unfortunately left a blank of at least .6 second between the removal of inspection figure and the presentation of the test figure. Regarding time taken to produce aftereffects after the fixation of an inspection figure and the time course of their decay, work such as that of Hammer (1949) has been performed but unfortunately no neurophysiological data is available which would permit direct comparison. A comparison between such times in figural aftereffects and simultaneous and successive contrast would permit us to test the relation expected on the theory.

It has been reported by Köhler and Wallach (1944) that outline figures have their maxima of effectiveness in displacing test contours at a greater distance than solid figures. Here it might be interesting to recall the situation occurring in *Limulus* and its theoretical explanation by Hartlone, Ratliff, and Miller (1961). Aggregates of excited cells exercise an inhibitory action on each other. A single ommatidium produces a higher frequency of discharge than when it is illuminated together with its neighbors. Similarly, the inhibitory effect of a whole group of excited receptors alters as its size is increased. The difference that would exist between line and solid figures is therefore not easy to predict without more precise information about contrast in the human visual system. However, here again, there should be a congruence between behavioral data on contrast and figural aftereffects.

However, it may here be objected that it is possible to obtain figural aftereffects with black lines on a white ground, and that on the theory derived from *Limulus* the black line would simply result in the cessation of firing of some receptors, and so only a slight reduction of inhibition in the illuminated receptors. The situation in mammals is more complicated. For instance, Baumgartner (1961) reports that in the cat one half of the investigated cells fire when not stimulated and then exercise a laterally inhibitory action on a large surrounding field. The other half are more similar to what we would expect from *Limulus*. These fire when they are illuminated and then inhibit a large surrounding field. Therefore, black lines could give rise to the type of lateral inhibition postulated here to account for figural aftereffects. Alternatively, the human system may be responding in terms of contours, rather than lines, as suggested by Köhler and Wallach (1944).

It might further be objected that the test contours would introduce their own patterns of inhibition and excitation into the situation as it is illustrated in Figure 2, which might produce all kinds of unexpected effects through processes of summation. This is true, but these new contours would be adding the same patterns of inhibition and excitation to the pre-existing pattern or baseline when there has been no previous prolonged inspection of other contours. Such patterns are presumably discounted, or any distortions of distance there are, are the distances we normally see. In the typical figural aftereffect experiment we compare a "normal" distance which has not been distorted by the introduction of previously fixated contours, but which still may be distorted in some way by the points we use to demarcate it (e.g., Oppel's illusion—Teuber, 1960), with a distance which has been distorted by our previous experimental manipulations. In other terms, the processes set up during the test are the same for the experimental and control stimuli and what the nervous system measures is the difference in the baselines for measuring distance on which both the control and the experimental stimuli ride.

However, it is worth noting that the theory would predict not only some illusions due to the presence of contrasting points during steady exposures, such as the filled-unfulfilled space illusion (Oppel, 1855, quoted by Teuber, 1960), but phenomena of increases in distance with progressively longer times of exposure. Ratliff (1961) and Hartline, Ratliff, and Miller (1961) report that the lateral inhibitory action of a stimulus lags its excitatory effect. Such a relation may be clearly seen in Figure 1 in comparing the consequence of stimulating one receptor. Its inhibitory action on another receptor (the first downward peak) is delayed in time in relation to the excitatory peak. We would expect, therefore, that when two points are flashed on, their distance should be underestimated and rapidly grow as the decrease in excitability in the space between them develops, due to an increasing lateral inhibition. This increase in apparent distance should then decrease with continued exposure. It is possible that some phenomena of apparent movement, such as the gamma phenomenon and underestimates of distance, when rapid rather than slower movement is induced, are due to this effect (Koffka, 1935; Teuber, 1960).

Further supporting evidence relating figural aftereffects to changes in neural excitability is contained in reductions of the threshold of pitch discrimination at a particular frequency after prolonged auditory stimulation at that frequency (Deutsch, 1951). Decreases in spontaneous activity in the auditory system following prolonged auditory stimulation have been found by Star[2] and Hilali and Whitfield (1953). If, as it has been suggested, discriminative acuity is related to the noise level in a system, then a decrease in threshold following prolonged stimulation can be attributed to a decreased noise level due to the preceding prolonged stimulation.

Earlier, we postponed a discussion of how increase or decrease in excitability could be translated by neural mechanism into distortions of distance between two points. Although the previously stated theory does not depend on our ability to state what such a mechanism is, the plausibility of the theory is increased if some suggestion can be made about the working of the postulated neural mechanism, so that what has been postulated can at least be seen to be possible and in principle relatively simple in its mode of operation. Two possible models will be suggested, but it should be stressed that the correctness of the hypothesis about figural aftereffects is independent of the speculations that follow.

In the first model, let us suppose that the distance between two points on a projection surface is estimated by the time it takes a wave to travel from one point to the other. Let us further suppose that the more excitable a region over which the wave travels, the more quickly it will travel over that region. This could occur, for instance, as a result of decreased synaptic delay in a more excitable region. Similarly, such a wave may be slowed down by a decrease in excitability in a given region. On such a system, if the region between two points were to become more excitable, then the time taken for the wave to travel between these points would diminish and so the apparent distance between the points would also diminish; the contrary would happen if the region became less excitable.

---

[2] A. Star, personal communication.

A second, entirely different system could give rise to the postulated relation between excitability and distance. We may suppose that two points are projected on a succession of superimposed sheets or planes. We may also assume that each plane produces an output when two points below a certain distance apart from each other are projected on it, thus acting as a summation device or a coincidence gate. Each successive plane is more sensitive to points further apart than the last so that points a great distance apart from each other will produce an output only in the last few planes. In this way, the distance apart of any two points will be estimated by which planes are producing an output. (For further details, see Deutsch, 1962.)

We can assume that if a portion of these planes is rendered more excitable, the plane will produce an output for two points lying in this portion, which would normally be only just too far apart to produce an output on this plane. As an output from this plane usually indicates the presence of two points which are closer, an underestimate of distance will be produced as a result of increased excitability between two points on such a plane. A similar argument can be used to show how an overestimate of distance would result from a decrease of excitability between two points.

CONCLUSIONS

The present suggestion attempts to reduce figural aftereffects to the same underlying physiological cause as the phenomena of simultaneous and successive contrast by making an assumption about how distances are computed by the nervous system. Since lateral inhibition occurs at all levels in the nervous system, the hypothesis does not postulate an exclusively central or peripheral origin of figural aftereffects. Before the hypothesis can be accepted, further work is needed examining the relation of contrast phenomena in their temporal and spatial aspects to figural aftereffects on the behavioral level; and on the more physiological level, work on mammalian nervous systems to investigate changes in activity at various distances from an excited locus after prolonged stimulation, when such stimulation has ceased.

REFERENCES

BAUMGARTNER, G. Corticale Mechanismen der Kontrast-Bewegungsehens: II. Receptive Felder. In R. Jung & H. Kornhuber (Eds.), *The visual system, neurophysiology and psychophysics.* Berlin: Springer-Verlag, 1961. Pp. 296–313.

BEVAN, W., JR. The influence of figural after-effects on visual intensity thresholds. *J. gen. Psychol.*, 1951, **45**, 189–207.

BROOKS, V. B. Contrast and stability in the nervous system. *Trans. N. Y. Acad. Sci.*, 1958, **21**, 387–394.

CRAIK, K. J. W., & ZANGWILL, O. L. Observations relating to the thresholds of a small figure within the contour of a closed-line figure. *Brit. J. Psychol.*, 1939, **30**, 139–150.

DEUTSCH, J. A. A preliminary report on a new auditory after-effect. *Quart. J. exp. Psychol.*, 1951, **3**, 43–46.

DEUTSCH, J. A. The statistical theory of figural after-effects and acuity. *Brit. J. Psychol.*, 1956, **47**, 208–215.

DEUTSCH, J. A. A system for shape recognition. *Psychol. Rev.*, 1962, **69**, 492–500.

Fox, B. H. Figural after-effects: Satiation and adaptation. *J. exp. Psychol.*, 1951, **42**, 317–326.

Fry, G. A., & Bartley, S. H. The effect of one border in the visual field upon the threshold of another. *Amer. J. Physiol.*, 1935, **112**, 414–421.

Gibson, J. J. Adaptation, after-effect and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1–31.

Hammer, E. R. Temporal factors in figural effects. *Amer. J. Psychol.*, 1949, **62**, 337–354.

Hartline, H. K., Ratliff, F., & Miller, W. H. Inhibitory interaction in the retina and its significance in vision. In E. Florey (Ed.), *Nervous inhibition.* New York: Pergamon Press, 1961. Pp. 241–284.

Hebb, D. O. *The organization of behavior: A neuropsychological theory.* New York: Wiley, 1949.

Heinemann, E. G., & Marill, T. Tilt adaptation in figural after-effects. *J. exp. Psychol.*, 1954, **48**, 468–472.

Hilali, S., Whitfield, I. C. Responses of the trapezoid body to acoustic stimulation with pure tones. *J. Physiol., London,* 1953, **122**, 158–171.

Hubel, D. H., & Wiesel, T. N. Receptor fields of single neurons in the cat's striate cortex. *J. Physiol., London,* 1959, **148**, 574–591.

Koffka, K. *Principles of gestalt psychology.* New York: Harcourt Brace, 1935.

Köhler, W. Relational determination in perception in cerebral mechanisms in behavior. In L. A. Jeffress (Ed.), *The Hixon symposium.* New York: Wiley, 1951.

Köhler, W., & Wallach, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 4.

Krauskopf, J. The magnitude of figural after-effects as a function of the duration of the test period. *Amer. J. Psychol.*, 1954, **67**, 684–690.

Kuffler, S. W. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.*, 1953, **16**, 37–68.

Mach, E. 1865, Über die Wirkung der räumlichen Vertheilung des Lichtreizes auf die Netzhaut. *Sitzber, Akad. Wiss.* Wien: Math. naturw. Kl. II, 52, 303–322.

McEwen, P. Figural after-effects. *Brit. J. Psychol.*, 1958, Monograph Suppl. 31.

Osgood, C. E., & Heyer, A. W. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

Ratliff, F. Inhibitory interaction and the detection and enhancement of contours. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Ch. 2.

Ratliff, F., & Hartline, H. K. The responses of limulus optic nerve fibers to patterns of illumination on the receptor mosaic. *J. gen. Physiol.*, 1959, **42**, 1241–1255.

Ratliff, F., & Riggs, L. A. Involuntary motions of the eye during monocular fixations. *J. exp. Psychol.*, 1950, **40**, 687–701.

Taylor, W. K. Visual organization. *Nature, London.* 1958, **182**, 29–31.

Teuber, H. L. Perception. In J. Field (Ed.), *Handbook of physiology: I. Neurophysiology.* Vol. 3. Washington, D. C.: American Physiological Society, 1960. Ch. 65.

(Received July 16, 1962)

Psychological Review
1964, Vol. 71, No. 1, 27-41

LONG-TERM MEMORY AS A FUNCTION OF RETENTION TIME AND OTHER CONDITIONS OF TRAINING AND RECALL[1]

EDWARD A. BILODEAU AND C. MICHAEL LEVY[2]

Tulane University

A general review of new methods, data, and recent interpretations of memory is made. Especial emphasis is given to the observation of 1000 Ss reported here and 1000 Ss examined earlier. It is found that motor and verbal methods of research have abruptly become more similar and have parted with the Ebbinghaus tradition in many ways. Rather than concentrating on between-subject phenomena, between-materials phenomena, and between-trials phenomena, the present approach related the events within-trials and also relates the events between 2 adjacent trials. Memory is described for 3 basic events: (a) response repetition, (b) response alternative, and (c) informative feedback. Changes in memory for these events are measured by changes in their interdependencies and upon their dependence on the prior events of Training. Many laws of retention are best specified by using indexes of consistency based on variance and correlation.

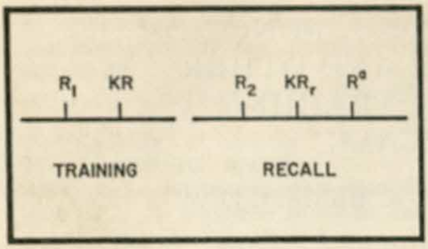
By 1958 much research in verbal learning gave the appearance of being fixed upon short-term memory and/or ancient dependent and independent variables. This was to change rapidly with Peterson and Peterson (1959), Underwood and Postman (1960), Sperling (1960), and Postman (1961, 1962) who have found limitations in old procedures, proposed new variables, introduced major new techniques, and altered our thinking on portions of the forgetting curve for verbal materials. An especially important procedural innovation, introduced by Peterson and Peterson (1959), was the use of a single training trial. Underwood (1957), in a brilliant analysis of forgetting, asked why the classical curve in verbal learning fell so far and so fast. He found, after making allowances for proaction, that verbal materials were very much less subject to forgetting than past data going back to Ebbinghaus (1885) had indicated.

Bilodeau and Bilodeau (1961) reviewed the contributions of the years 1945–1959 to the area of forgetting of motor skills and showed once again that most studies of such forgetting failed to find it. They were unable to find any full-scale retention curve—a failure that, of course, made it impossible to summarize the variables determining motor skills forgetting and their roles in producing more or less of it. By 1962 there were only portions of curves in verbal learning and still no curve to speak of in motor learning. The primary object of this review is to examine the writers' full-scale, long-term retention curves and to discuss the critical variables controlling forgetting and failure of forgetting of skills.

---

[1] This work was financed by Grant G-13294 of the National Science Foundation.

[2] The writers are much indebted to J. L. Sulzer for his many contributions, to Audrey Ayo and William H. Saufley, Jr. for their skillful assistance, and to Ina McD. Bilodeau for her help with the manuscript.



CRITICAL EVENTS OF MEMORY

FIG. 1. Identification of the two major events of a trial of Training and the three major events of a trial of Recall.

A Remedy

In a monograph on motor and verbal memory, Bilodeau, Sulzer, and Levy (1962) have asked what primary events might be subject to forgetting. Their analysis centered on the analysis of components within a single trial of training and a single trial of memory and was recommended for both verbal and motor applications. As a matter of fact, the verbal/motor dichotomy was not needed except to specify training materials. The critical events of Training and Recall are shown in Figure 1 and are defined as follows. The R_1 is the first and only response of Training, the R^a is the alternative response to R_1, and the KR is the informative feedback pertaining to R_1 or the feedback pertaining to R^a. In Recall, R_2, R^a, and KR_r are the subject's attempts at recollecting the earlier events. The R_2 represents the subject's attempt to repeat R_1; R_2 is not an attempt to improve upon R_1. Components KR and KR_r, such as "You were low by 54 units," "You told me I was high," are more verbal than motor and emphasize the early involvement of verbal matters even at the roots of motor responding.

The response terms R_1, R_2, and R^a may be either motor or verbal depending upon the experiment the experimenter chooses to do. The R_1 and KR are the two observable events of Training. The R_2, KR_r, and R^a are the recallable events of the very next trial. An important point to bear in mind is that R_1, R^a, and KR can be made variable, continuously scaled quantities as in a lever positioning task. In the remainder of this paper the events are treated as continuously variable even though they are not always; i.e., when the subject repeats a training word or when the experimenter says "correct." The object of using scaled variables was to relate the individual events of the recall trial to those of the training trial and to each other more precisely. The literature contains little about these relationships and, indeed, even discerning the individual events *within trials* is blurred. Certainly, when we do separate the within-trial events, it is clear that more than further investigations of phenomena between subjects, between materials, and between trials, we require a series of studies relating the events within trials and the events between adjacent trials. It was thought that the task of relating could best be begun by examining consistencies and changing consistencies between the primary events. For example, it was soon found that with increasing retention intervals, forgetting the events in a lever-positioning task showed in increasing variances, decreasing relationships (r^2) between the events of Recall and Training, and increasing relationships (r^2) between the events of Recall.

The analysis of component events suggested by Figure 1 led to a great number of questions which could not be answered without getting into many particulars. To take just one example: will the subject's long-term recall of R^a depend more upon (*a*) the recalled KR, or (*b*) the administered KR? The memory monograph contains the complete formulation and considerable amounts of data as well. The present study deals mainly with the effects of the retention interval, an interval ranging up to 6 weeks where no special interpolated activity is required of the subject and where normal daily activities obtain.

One of the effects of the previous study was to show the usefulness of variance and correlation in measuring retention. This is not at all surprising for it is well established that the variable error contains information which a measure of central tendency does not. And these measures are especially important when the amplitude of response represents the initial observation. In the past, most analyses of memory could be reduced to the form: there is no difference between the mean frequency of items, mean time on target, etc., of Training and Recall. Yet means can remain constant while major changes occur in the scores from one distribution to another. Individuals will begin to stray farther from the mean and to exchange position within the distribution.[3] These changes in the data are often indications that retention is changing. Actually, there are seven ways to describe potential changes in memory.

Measures of Memory

Changes between two arrays of scores (Training versus Recall) can be measured by M, σ^2, and r. Either one, two, or three of these statistics can change from Training to Recall in a total of seven combinations. Textbooks in the psychology of learning ordinarily reduce these seven to one and perhaps in the long run to the most trivial of the seven, i.e., the average amount of material recalled is less than the average previously learned. Modern theories of learning are silent on the question of the environments that will produce one or another of the seven dependent possibilities. The writers do not know either.

Figure 2 has been drawn to encourage further questions about variance and correlational phenomena (and any other statistic of change) in memory research. In Part A, R_1 represents the distribution of the first response of a large number of subjects. The distribution R_2 represents the subjects' attempts to reproduce their previous R_1. The arrows show what might happen to four hypothetical subjects. In short, they do so badly at recalling (R_2) that the correlation between the two arrays of scores ($R_1 \cdot R_2$) falls to zero and reflects what is happening within subjects from one trial to the next. Yet, by comparing the two distributions still further we see that the means and variances hold constant. The mean and variance have been completely insensitive to the very substantial amounts of forgetting taking place. Part B of Figure 2 represents the case of change of variance, no change in mean, and no change in correlation. In this case of forgetting, the subjects can be said to be getting wilder than ever, yet Z scores and means remain the same from Training to Recall.

In the experiment reported below we look for evidence on the form of the retention curves for R_2, KR_r, and R^a as measured by variance and correlation under conditions of simple training and recall environments. The aim is to consider memory as an aspect of response consistency and to consider the merits of the hypothesis that forgetting is well expressed in terms of response inconsistency.

---

[3] Lewis (1959) has made serious use of r in studies of retroaction of the OL-IL-RL design.

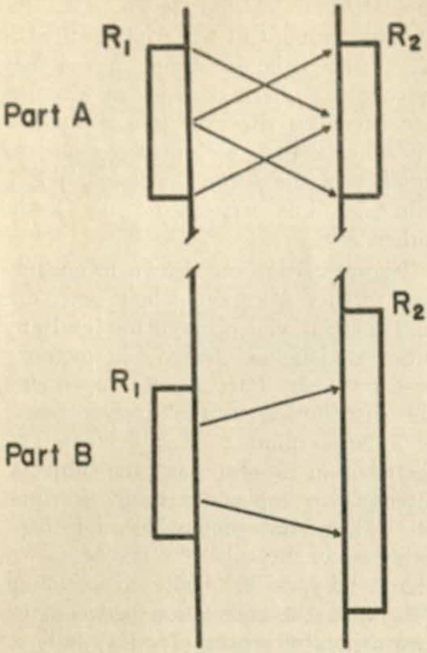


FIG. 2. In Part A, R_2 represents a distribution of responses where the subjects attempt to repeat their prior responses (R_1). Here, forgetting has led to a correlation of zero, but mean and variance have held constant. In Part B, forgetting has led to an increase in variance, but mean and correlation have held constant.

METHOD

Subjects and apparatus. Over 1000 subjects were trained and 946 of these were tested for recall. Half were scheduled from college sources and half from Keesler AFB. However, since absences were greater in the colleges, the colleges ultimately yielded 380 subjects and Keesler produced the remainder.[4] The proportion of college subjects from group to group was fairly constant. Furthermore, the college and airmen subjects have behaved alike during the present series of studies and mixing these subjects is now routine.

The apparatus was Manual Lever D, a box housing a movable lever (Sulzer, 1963). The subject's task is to stand in front of the box, grasp the unseen lever, and move it from left to right on the command of "move." Two targets, one corresponding to R_1 and one to R^a, were employed. Target 100 was 20° beyond the response origin and 64° above the horizontal. Target 300 was 60° beyond the origin. Responses and Targets may be converted into degrees of arc by dividing by 5.0.

Procedure. During *Pretraining,* the subject moved the lever (twice) until it stopped automatically at one, then the other, of the two targets. Thereafter, the experimenter removed the automatic stops and the subject had to stop the lever without any mechanical assistance. During *Training,* the subject was required to guess which of the two targets was "correct" and to indicate his selection by executing a move (R_1) blindly and without the aid of the stop. His KR was administered immediately after R_1. The procedures used were almost exactly those described for Group I-28 in the earlier study.[5] *Recall* followed the R_1 and KR events by 3 minutes, 20 minutes, 2 days, or 6 weeks. These retention intervals will be used to designate the four groups.

The groups were subdivided according to the amplitude of R_1 and the KR administered as follows:

1. 100 Shooters: subjects for whom $R_1 < 200$. This defines their alternative target as 300 and their KR as Low or Correct.

---

[4] The permission of the Commander of Keesler Air Force Base and the assistance of Major Marvin Karp are gratefully acknowledged.

Civilian subjects were obtained from Louisiana State University, Baton Rouge; Louisiana State University, New Orleans; and Tulane University, New Orleans.

[5] The experiments are highly related and were designed so that their data can be compared. This also means that the elements of minor detail in common need not be rewritten here. The most important departure from the previous study was the omission of the probes for R_3 and R^a_2 (the subject was not asked to repeat the R_2 and the R^a he had produced a few seconds before). It had been shown that the reliability of the present task is very high, for the correlations between responses separated by a few seconds are over .90. There was no need to show this again and valuable time was saved by dropping the two probes. The effect upon the subject was probably to make recall a trifle easier in that R_3 or R^a_2 constitutes an extra interpolated task.

(*a*) Low: "You chose the wrong target and did not move far enough by —— units."
(*b*) Correct: "You chose the correct target."

2. 300 Shooters: subjects for whom $R_1 > 200$. This defines their alternative target as 100 and their KR as High or Correct.

(*a*) High: "You chose the wrong target and moved too far by —— units."
(*b*) Correct: "You chose the correct target."

It might be helpful to summarize the task by following a hypothetical subject. For example, if R_1 equals 115 units, the subject is a 100 Shooter by Definition 1 above and has overshot the target of his choice by 15 units. Further, for KR the experimenter might say, "You chose the wrong target and did not move far enough by —— units" (1*a* above), and dismiss the subject. After 6 weeks, in response to the experimenter's request to repeat R_1, the subject might move 130 units instead of 115; in response to the experimenter's request for R^a, the subject might move 310 units instead of 300; in response to the experimenter's question about the previous KR, the subject might answer, "You said I did not move far enough by about 50 units." Now that the procedure has been made so specific it would be well to relate it to Figure 1. The R_1s, R_2s, and R^as are simple positioning responses; KR and KR_r are verbal events; a request for R_2 (or R^a) illustrates a selective learning component in that the subject must choose between classes R_2 and R^a. The foregoing analysis makes it impossible to describe the present task as motor, verbal, or verbal-motor. Instead, it is all of these.

A unique use of KR prevailed, and the logic of it is described fully in the monograph. Nevertheless, to reduce a potential source of confusion, a brief overview of the KR procedure seems in order. The experimenter's choice of values for KR was arbitrary in order to make KR independent of R_1. This makes it possible to relate the events of Recall to two independent events of Training. The subjects were prescheduled to be Correct or to be given KR of a particular magnitude if Incorrect. The magnitude of R_1 served to determine if the subject were to be Low or High, but the size of R_1 did not determine the size of KR. The values of KR were determined by the experimenter and by sampling rectilinearly from the series 27, 54, 81, . . . 324. Thus, as both the number and the unit were arbitrary, it was expected that such KRs would harm the recall of R_1 and R^a as contrasted with the customary expectation of improving responses through feedback.

The subject's activities during the retention intervals must be made explicit, for these activities may produce some of the failures of memory to be reported. Precautions were taken to prevent the subject from realizing that he was being trained and that the experimenter would return to test his recollection of R_1, R^a, and KR. The subjects of Groups 2 Day and 6 Week were dismissed after Training and the experimenter returned to them in their classroom at the time for Recall. There was, however, so little time between Training and Recall for Group 3 Minute that none of the subjects was sent back to class. This was also the practice for Group 20 Minute, except for the college subjects, who through necessity had to be returned to their classroom lectures during their layover. It cannot be said that the four retention groups have had identical experiences during their intervals. It was and remains technically impossible to arrange homogeneous retention environments for all four groups and we are not convinced that this was an important shortcoming. It may never be possible to have the content of a 6-week interval qualitatively and quantitatively the same as for one of 3 minutes. This is a type of limitation, but not a serious one in view of the main objective of examining and understanding the kind of recall behaviors that follow such normally occupied intervals of retention.

Moreover, the experimenter used two procedures to deliberately disengage the subject from the immediate training environment and, thus, to put all groups at least on the same continuum. First, every subject broke the position and posture assumed for the training trial; even the 3-Minute subject was dismissed from the room and training environment (for 2 minutes or so). Second, every subject interpolated other activity as he walked and talked during the interval. Finally, all groups were reinstructed on position, posture, grip, and commands just before the beginning of Recall.

RESULTS

The distribution of the 946 subjects is shown as the lower entry in the cells

TABLE 1

NUMBER OF SUBJECTS REVERSING TARGETS AND TOTAL SUBJECTS

R_1	KR	Group			
		3 Minute	20 Minute	2 Day	6 Week
100 Shooters	Low	3/73	0/70	2/65	9/52
	Correct	1/59	1/50	0/49	5/44
300 Shooters	High	2/67	5/63	6/73	9/65
	Correct	2/53	1/58	2/58	8/47

of Table 1. The upper entries show the numbers of subjects who selected the wrong target or who reversed the two target places during Recall. For example, after 6 weeks, 9 of 52 Low subjects responded with R^a when asked for R_2 and responded with R_2 when asked for R^a. Reversal is a rather infrequent type of forgetting, but especially so at 3 and 20 minutes. Even though the relative frequencies in Table 1 never exceed one in five, the reversing subjects have a powerful effect upon the statistics which follow. Reversing subjects fall outside the main distribution and contribute more than their share to M, σ^2, and r. Unfortunately, the worst hob is raised at a critical spot, among the Low of the 3-Minute 100 Shooters, where an upward fluctuation produced three reversing subjects, and in Group 20-Minute 100 Shooters Low, where a downward fluctuation produced no reversing subjects. As we shall see, the overall effect of these two fluctuations was to produce inversions of rank order between Groups 3 Minute and 20 Minute and to give the appearance of better recall with the longer interval in the 100-Shooter analyses which follow.

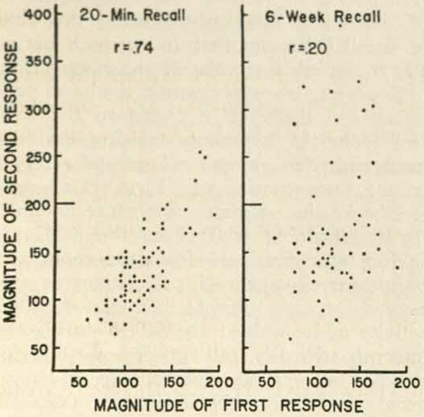


FIG. 3. Scatter plots for 100 Shooters showing the deterioration of R_2 with respect to R_1 after 20 minutes and after 6 weeks.

Figure 3 was prepared as a scatter plot to give the reader a better visualization of the forgetting analyses to follow. The R_2 is plotted against R_1 for Group 20 Minute on the left side and Group 6 Week on the right side. Group 6 Week shows a greater variance in R_2 and a lesser correlation for $R_1 \cdot R_2$ than Group 20 Minute. The inconsistency of recall in Group 6 Week is further heightened by the fact that 9 of 11 subjects with $R_2 > 200$ executed $R_2 > R^a$.

The data of the earlier study made it clear that Target 300 presents an easier task than does Target 100. It is for this reason that the analyses below emphasize Target 100 over 300.

Effects of Forgetting upon the Means

The means for the 100 and 300 Shooters are given in Tables 2 and 3, respectively. The 100 Shooters show the inversion between the first two groups mentioned before, and then the means of R_2 deteriorate progressively (toward 300) with intervals of 20 minutes, 2 days, and 6 weeks. The mean of 177 for Group 6 Week is very much the same as obtained in the earlier study with a 4-week interval, but the mean of 159 for Group 3 Minute is much higher than expected. The R_2s at the bottom of Table 2 are for the Correct subjects and are more in line with expectations. Furthermore, in all four cases the Correct subjects have outperformed the Low subjects ($p < .05$, $df = 1/454$) in the attempt to repeat the original response.

The R^as for the 100 Shooters show little forgetting taking place at Target 300. The greatest evidence for forgetting is the 292 of Group 6 Week Low, which, though quite an accurate R^a, is a deterioration from a 330 level which would have resulted if the test had been scheduled earlier. There is a significant difference at the .05 level between the Low and the Correct subjects at 6 weeks ($t = 2.07$, $df = 94$).

TABLE 2

MEANS FOR THE 100-TARGET SHOOTERS

	Group			
	3 Minute	20 Minute	2 Day	6 Week
Low				
KR	−165	−181	−165	−183
KR_r	-133^1	-139^2	-100^3	-31^4
R_1	127	118	129	122
R_2	159	133	156	177
R^a	327	327	347	292
Correct				
R_1	124	118	128	124
R_2	134	125	142	175
R^a	333	316	339	334

Note.—Superscripts 1, 2, 3, and 4 are:
[1] $n = 72$ of a possible 73.
[2] $n = 69$ of a possible 70.
[3] $n = 62$ of a possible 65.
[4] $n = 44$ of a possible 52.

TABLE 3

MEANS FOR THE 300-TARGET SHOOTERS

	Group			
	3 Minute	20 Minute	2 Day	6 Week
High				
KR	185	177	189	168
KR_r	154	162^1	152^2	71^3
R_1	315	308	320	325
R_2	300	298	305	295
R^a	133	134	152	181
Correct				
R_1	315	311	321	325
R_2	310	306	315	295
R^a	132	129	140	196

Note.—Superscripts 1, 2, and 3 are:
[1] $n = 60$ of a possible 63.
[2] $n = 72$ of a possible 73.
[3] $n = 56$ of a possible 65.

Examination of the means of the 300 Shooters in Table 3 will show no sign of forgetting in R_2 and strong indications of forgetting in R^a. The Correct subjects perform no worse at R^a than the High, and the F between feedbacks is less than unity ($df = 1/476$). At 6 weeks neither the High nor Correct performs well, for with scores of 181 and 196 they are approaching a point midway between targets. These values do not differ from one another by a t test.

The Correct subjects do exceptionally well at recalling that they were told "correct." Of 202 correct subjects moving toward the 100 Target, 2 did not recall their KR (both from Group 6 Week). The Low subjects have more to forget, it being possible to forget that they were *wrong, low* and *x units low.* Of the 260 subjects in

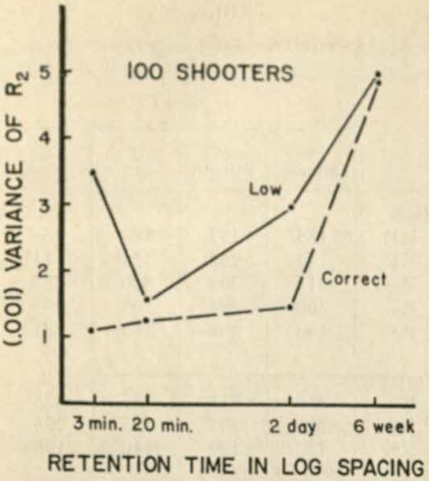


FIG. 4. An illustration of increasing variance in a repeated response as a function of increasing value of retention time. (The state of KR serves as a parameter.)

the Low category, 247 recalled some number and some sign. Eight of the 13 subjects who did not were in the 6-Week group and responded (*a*) with nothing, (*b*) "I was wrong," or (*c*) "I was —— off." The data of the 300 Shooters were very similar. The category of subjects recalling some number and some sign is represented as KR_r in Tables 2 and 3. All four groups show the tendency of recalling a figure smaller than the one administered. The mean for Group 6 Week is −31 and is smaller in absolute terms than the −183 administered. However, 15 of the 44 subjects are mistaken about the sign of KR and 5 of these confuse R_2 and R^a as well. Their influence on the means of Table 2 is great.

Effects of Forgetting upon the Variances

The variance of the recalled event serves as the dependent variable and the length of the retention interval serves as the independent variable.[6] The four intervals were converted to minutes and then to logarithms. Otherwise, it would have been technically impossible to express the distances between the various rests (ranging from 3 to over 60,000 minutes) on graph paper. The log scale, moreover, does no violence to the shape of the curves to be reported.[7] The log scale has the further virtue of grouping the first two points and grouping the last two points. This makes it easier to determine whether the curves are negatively or positively accelerated inasmuch as two points of inquiry are placed at both ends of the scale.

Figure 4 shows the variances for R_2 of the 100 Shooters. The curve with the lesser variances was produced by the subjects told that R_1 was Correct. The curve for Correct is clearly positively accelerated while the acceleration for the Low is not so clear because of the (unreliable) location of the first point. The writers would have been pleased to have the three subjects who responded as if R_2 were R^a somewhere else, e.g., in a void pile. Making an allowance for the three subjects would produce a curve resembling that for Correct. A subsequent experiment confirmed the hypothesis that the shape for Low resembles the one for Correct.

The R_2 variances for Target 300 are not presented because recalling

---

[6] Tables of variances and correlations for each of the groups have been deposited with the American Documentation Institute. Order Document No. 7620 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance $1.25 for microfilm of $1.25 for photocopies, and make checks payable to: Chief, Photoduplication Service, Library of Congress.

[7] No transformation of the abscissa can alter the fact that the first two or three ordinal values remain constant.

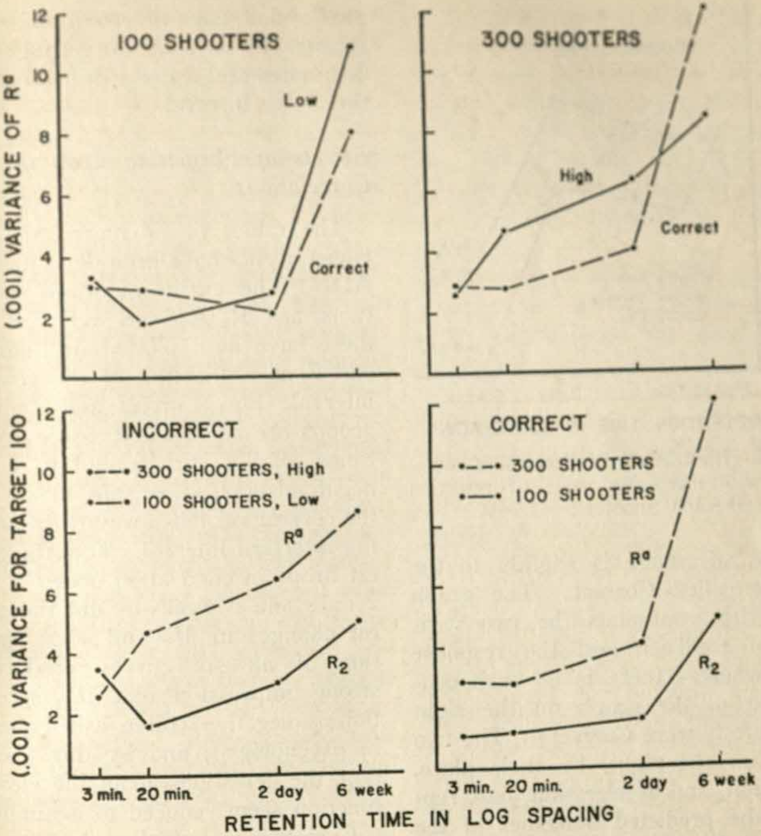


FIG. 5. Four illustrations of increasing variance in the recalled response as functions of the conditions of feedback and responses required.

this response, as in earlier studies, is so easy. The variances for the first three groups were quite the same and the ratios of R_2/R_1 were close to unity. At 6 weeks, however, forgetting was substantial and the variance for R_2 was nearly four times as large as for R_1. Further, the Highs and the Corrects were very much the same at each interval.

The variances of R^a in the top two graphs of Figure 5 are for 100 Shooters and 300 Shooters. Note the strong tendency for positive acceleration in three of the four curves. Further, there is no substantial difference between the Low and Correct and between the High and Correct. However, there is a strong tendency for the variances of R^a to swell more in the 300 Shooters than in the 100 Shooters because they have the harder task to remember. The reader should recall that Target 100 produces a slightly smaller R_1 variance than does Target 300.

The pair of figures at the bottom of Figure 5 compares the recall of R_2 and R^a *with target place held constant* at Target 100. The one on the left is for KR Incorrect and the one on the right is for KR Correct. The results make good sense in that R^a is more poorly recalled than R_2

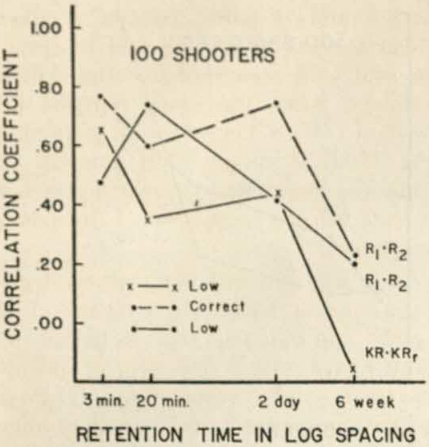


Fig. 6. Illustrations of decreasing correlations with increasing value of retention time for the 100 Shooters.

and the advantage is slightly to the subjects called Correct. The graph on the left confounds the two variables of feedback and the response factor whereas there is no such confounding in the graph on the right (all subjects were Correct). The two curves on the right, by their place, interaction, and acceleration, best represent the predicted outcomes of the experiment.

Figures for Target 300 are not presented. There was no difference in the recall of R_2 and R^a for the first three retention intervals. At 6 weeks, however, R^a deteriorated markedly for both the Low and the Correct comparisons. The deterioration was worse for Low than for Correct because of the stimulus properties of KR.

Lastly, it should be stressed that in the preceding analyses of means and variances, the data showing the greater forgetting have been made into graphs, whereas the data with lesser forgetting have been summarized in words. All the data are consistent with the statement that little or no forgetting takes place *between* 3 minutes and 2 days of retention. There appears to be some forgetting before 3 minutes and a great deal more after the 2-day interval.

Effects of Forgetting upon the Correlations

The tables of intercorrelations between events have been deposited with ADI. The extra important correlations are $R_1 \cdot R_2$ and $KR \cdot KR_r$ and these have been plotted in Figures 6 and 7 as a function of the retention interval. Low, Correct, and High subgroups are represented as parameters. Since all six curves tend to have a negative slope it is safe to say that the correlation falls away with increasing intertrial interval. Too, the greatest drops in correlation occur between 2 days and 6 weeks as did the greatest changes in Ms and σ^2s. Inspection of all six curves produces a strong impression that the acceleration is negative (often called positive in psychology), and by this assumption, the deviations about the idealized function seem reduced to a minimum.

Correlation coefficients are notori-

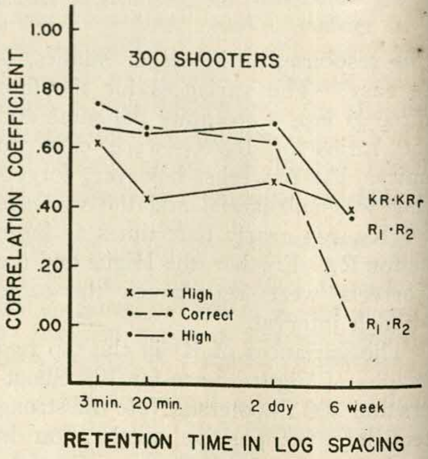


Fig. 7. Illustrations of decreasing correlations with increasing value of retention time for the 300 Shooters.

ously unreliable and so comparisons between pairs of curves are made at a considerable risk. We prefer not to compare them. Because of inversions and other considerations the writers find insufficient evidence to establish a difference between Correct and Low and between Correct and High. There is evidence in the 100 Shooters that the inconsistency within the distribution of KR_r is greater than that for R_2, but this is not fundamental as it depends on the memorability of the particular KR numbers used. Moreover, the $KR \cdot KR_r$ curve is already purged of certain subjects who fail to recall KR. When comparing the 100 Shooters against the 300 Shooters the most noticeable effect is the greater fluctuation in the r's of the 100 Shooters. It is hard to decide whether this is a property of the target place per se or whether it is attributable to the fact that *some* subjects find it difficult to recall R_2 at Target 100. The preceding difficulties, however, do not bear on the conclusion that the three curves suggest a negative acceleration.

DISCUSSION

A search of the literature on memory for motor skills by Bilodeau and Bilodeau (1961) found no full-scale forgetting curve worthy of the name. They also concluded that interference has been difficult to produce when material is deliberately interpolated between original learning and relearning. A great deal of forgetting, however, was shown in the data of the present series. Why forgetting was so strong here and so weak before is a question that brings up several systematic issues and important leads.

1. The present studies use a finely grained measure of response—amplitude in degrees of arc. Oftentimes, grosser measures such as time-on-target and response frequency allow diverse ways of scoring and it cannot be said with assurance that the skill is retained when the score remains invariant. It is important to note that the kind of motor skill that is so famously resistant to forgetting is not a rote-learned sequence of invariant responses.

2. Variances and correlations were examined as well as the mean. For some applications, the statistics σ^2 and r are more sensitive to inconsistencies of responding. Lewis (1959) has used correlation in studies of the effect of interpolated learning upon retention, but few others have had adequate size of sample or the inclination. Bilodeau et al. (1962) have successfully employed and described what happens to the means as forgetting of the present task takes place, but they have also urged the wider use of indices of change within individuals. In motor research, error or magnitude of response is observed as a matter of course, and variance and correlation statistics naturally follow.[8]

3. The amount of training and material were both severely limited. During Training R_1 was executed only once and the experimenter did not ask for R^a. The KR administered was a single stimulus event to which the subject made no observable response during Training. Training was further minimized because the subject was not told that he was being trained nor that he would later be asked for his recollection of anything. The pretraining and training

---

[8] Criticism of our urging wider examination of changes of the subject's behavior between Training and Recall has taken the form that many tasks are not amenable to variance or correlational analyses as suggested in this paper. The critics are correct; many tasks are not.

given were adequate to produce high levels of intertrial consistency for periods up to a day or so, but were insufficient to last 6 weeks.

Reviewing the years of research on rote learning of verbal materials, Underwood (1957) stated that future experiments should stress the learning of a single association and the measurement of its decline in strength over time. By this he meant one training trial on one response item. The present study stresses single presentation too, but it uses *three* primary events. Even simpler experiments than the one reported here are possible. Future designs could eliminate R^a and KR and employ only R_1, R_2. On the other hand, the three events could be retained, and the source and level of recall controlled within the individual by reminding the subject of one or two events before asking for the recall of the other.[9] The retrieval of an event should depend on the state of recall of another (or the presence of another) and so the important question becomes the degree to which they are associated either through pre-experimental means or through training.

Peterson and Peterson (1959) have measured the short-term retention of individual items (R_1, R_2) in a very original study. There remains, however, the necessity of determining the retention of R^a and KR in a more verbal context than levering. The analysis of KR, in particular, has been consistently overlooked in verbal learning because KR is often confounded with the response term. In lists of paired associates, for example, $x - y$ is exposed after the stimulus term x, y being ordinarily considered the response term. Indeed, though it has been noted before that y serves as a confirmation as well, its KR function has in fact been neglected. A simple analogy to the present experiment in a verbal context may make the point that verbal and motor retention can in fact be studied at the same level. For example, a verbal response might be the word selected from a number of alternatives, as in multiple choice. If the subject were shown Items X, Y, Z simultaneously, and chose Z, we could observe the forgetting of the R^as (X and Y) and the KR (correct or incorrect), and their interdependencies. For further analysis of verbal experiments in terms of R^1, R^a, KR see Bilodeau et al. (1962).

4. Most previous motor skills studies have used tracking tasks where both stimulus and response are continuous and complex. Thus, we are in the position of obtaining more forgetting with a simple task and little training than with a complex task and considerable training. The complex task situations seem to have produced high intratask transfer. Item 5 below bears on this perplexity.

5. A long-standing enigma is that verbal skills are rapidly forgotten while motor skills are exceedingly well retained. Much of this problem has now disappeared because (*a*) Underwood (1957; see also Keppel & Underwood, 1962) showed that forgetting of verbal material is not necessarily rapid and great, and (*b*) the present studies show that the forgetting of a positioning response, such as R_1 or R^a, can be great. Recent years have brought changes in our expectation of the shape of the retention curve for both motor and verbal material. The classical (Ebbinghaus) curve has lost its pre-eminent position. Other techniques are producing curves with

---

[9] Work on long-term retention under conditions of artificially induced recall has been carried out by the writers (Bilodeau, Levy, & Sulzer, 1963). Stimulated recall bears on modified free recall, whereas the present paper deals with free recall.

other shapes. Even the immediate memory span has been reinvestigated, and this has led to forgetting functions within the "span" (Sperling, 1960). To claim a comprehensive theory of human retention would be premature. But it is not overly optimistic to say that recent treatments of both verbal and motor retention represent advances toward an account of human retention and its controlling variables. The cheerfulness of the present outlook depends on the observation that *both* verbal and motor forgetting can, with proper control, deviate from their stereotypes. It has become possible to talk about both in the same article, reversing the trend of several decades.

Reanalyzing classical pre-World War II data, Underwood reported that 85% or more of the items are generally retained over 24 hours, and says, "In short, we must start all over and derive a retention curve over time when the subjects have learned no previous materials in the laboratory. It is apparent that I expect the fall in this curve over time to be relatively small." The shape of the curves for levering is well in line with Underwood's expectation for we did find little evidence of forgetting between 3 minutes and 2 days of retention.

6. The writers suspect the existence of a two-stage curve of forgetting for positioning tasks of the lever variety. The subjects of all four of our groups disengaged themselves from the lever and training environment between Training and Recall. That is, by removing themselves from the training situation, all subjects did such things as: released the grip of the lever, moved the resting arm away from the top of the leverbox, changed foot position, moved out of the room, moved back, etc. All of these events occurred between R_1 and R_2 and the highest $R_1 \cdot R_2$ correlation found was in the vicinity of .70. However, in the study by Bilodeau et al. (1962), some subjects were left engaged between successive responses and only a few seconds intervened between R_1 and R_2. These $R_1 \cdot R_2$ correlations were always close to .95. Therefore, it seems that the act of moving away from the lever after R_1 drops the correlation from .95 to .70 (and nearly doubles the variance) and suggests why retention was less than perfect at the 3-minute point. Further inquiry is certainly required, but we believe there might be two decay stages to a full-scale forgetting function: a brief and shallow, but very rapid, first stage followed by a relatively long and accelerating second stage as shown in Figure 8. The evidence for the hypothesized second stage (solid line) is to be found in the section on Results; the first stage (dashed portion at left) has yet to be studied in detail with the lever. A virtue of Figure 8 is that the Y intercept is fixed at 1.00. Here, recall is perfect in the sense that the variance ratio of R_2/R_1 is unity or the correlation for $R_1 \cdot R_2$ is unity.

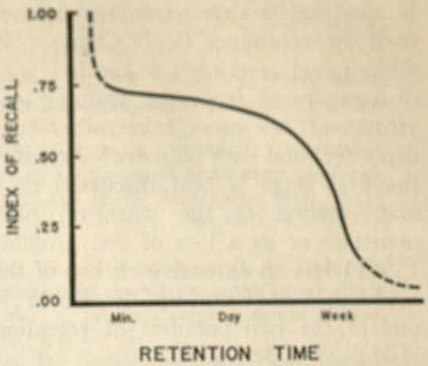


FIG. 8. Hypothesized two-stage forgetting curve.

Disengaging the subject could produce (*a*) responses which interfere with the recall of R_1 or (*b*) changes in postural and environmental cues used to reproduce R_1. Changes in stimuli and responses associated with disengagement from the training environment, we guess, take rather sudden effect and then no more. Actually, the first stage is best discussed as a manifestation of the warm-up phenomenon or as a loss of set. Adams (1961) has an extensive review of the evidence that warm-up decrement is one of the two portions of retention loss.

The second stage represents a grosser type of forgetting—the one studied in this series. We need to know more about the explanation of the second stage and more about the description of the first stage. We guess that the first stage is completed within minutes or less, that the near-plateau period lasts for a day or two, and that the greatest amount of forgetting occurs after a few days and before several weeks. Some past studies of motor forgetting may have placed too many of their empirical points between the two stages of decay and thus found no forgetting.

The acceleration of the second stage is novel, since the Ebbinghaus curve shows the opposite sign. Data supporting the Ebbinghaus function are ordinarily explained by referring to exponential-like decay processes—a popular explanation because so many kinds of data can be fitted by the exponential that it might almost seem the only natural process. Using the idea of reliability, however, provides an attractive way of rationalizing the unusual acceleration of the second stage of Figure 8. Given that the subject can repeat a response from one moment to the next, he should also be able to repeat nearly as well if the attempted repetition is delayed by an extra moment. Said another way, if the probability of repeating a response exactly over interval t is very high, there is no reason to assume that something unusual will happen over $t + \mathrm{k}$ (where t and k are of the same order of magnitude). By analogy, assume that a car is started once a day and that there are no failures at starting. What is the probability of failure with the interpolation of an extra day? Somewhat greater than before to be sure, but the writers doubt that the greatest increment in probability occurs in the first (of many) extra day of layover. Only a trifle of updating of pro- and retroaction concepts is required to explain the delay of forgetting shown in the second stage of Figure 8. The updating is based on the general idea that the probability of the occurrence of interfering events should rise in keeping with the expectation of events randomly distributed in time and that the growth of their effect is ogival.

7. Last but not least, it seems reasonable to suppose that using two targets was responsible for some of the forgetting shown here. It was the use of a second Pretraining target which established R^a. We have seen R^a produced when the experimenter asked for R_2, and we have seen the difference between the means for R^a and R_2 approach zero with increasing retention intervals. The two target habits appear to interfere with each other and to do so increasingly with time. In the end there could be complete forgetting of class membership, an issue analogous to list membership (first or second) as discussed in verbal learning (Barnes & Underwood, 1959; Postman, 1961).

Only two targets have been used with the lever, though more could have been used; since one of our concerns is to produce motor forgetting new studies should include more R^as. The work of Reid, Lloyd, Brackett, and Hawkins (1961) and of Murdock (1961), who varied the number of verbal items constituting the stimulus, is relevant. In the levering series, there were two types of loads: (*a*) one, two, or three basic events (R_1, R^a, KR) and (*b*) number of items within events (held at unity). Variations of load will quickly move the inquiry to considerations of proaction because learning and memory should succumb with increasing load. And memory should also tend to fail as a function of responding qua responding (retrieving) during Recall.

SUMMARY

Motor and verbal methods of research, forgetting phenomena, and explanations have abruptly become more similar, reversing a long apartheid-like trend. It seems desirable to reference motor and verbal research in the same article, perhaps because both approaches speak to human memory. The closer ties have also meant substantial departures from the school of Ebbinghaus.

REFERENCES

ADAMS, J. A. The second facet of forgetting: A review of warm-up decrement. *Psychol. Bull.*, 1961, **58**, 257–273.

BARNES, J. M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

BILODEAU, E. A., & BILODEAU, I. McD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, **12**, 243–280.

BILODEAU, E. A., LEVY, C. M., & SULZER, J. L. Long-term retention under conditions of artificially induced recall of related events. *Percept. mot. Skills*, 1963, **16**, 875–910. *Monogr. Suppl.* 8–V16.

BILODEAU, E. A., SULZER, J. L., & LEVY, C. M. Theory and data on the interrelationships of three factors of memory. *Psychol. Monogr.*, 1962, **76**(20, Whole No. 539).

EBBINGHAUS, H. *Ober das Gedächtnis: Untersuchungen zur experimentellen Psychologie.* Leipzig: Duncker and Humblot, 1885. (Trans. as *Memory: A contribution to experimental psychology* by H. A. Ruger and C. E. Bussenius. New York: Teachers College, Columbia University, 1913.)

KEPPEL, G., & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single items. *J. verbal Learning verbal Behav.*, 1962, **1**, 153–161.

LEWIS, D. Latest news and views regarding individual susceptibility to interference in the performance of perceptual-motor tasks. Presented at Midwestern Psychological Association, Chicago, Illinois, 1959.

MURDOCK, B. B., JR. Short-term retention of single paired associates. *Psychol. Rep.*, 1961, **8**, 280.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961. Pp. 152–196.

POSTMAN, L. The temporal course of proactive inhibition for serial lists. *J. exp. Psychol.*, 1962, **63**, 361–369.

REID, L. S., LLOYD, K. E., BRACKETT, H. R., & HAWKINS, W. F. Short-term retention as a function of average storage load and average load reduction. *J. exp. Psychol.*, 1961, **62**, 518–522.

SPERLING, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

SULZER, J. L. Manual Lever D: A basic psychomotor apparatus for the study of feedback. *Percept. mot. Skills*, 1963, **16**, 859–863.

UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

(Received August 6, 1962)

Psychological Review
1964, Vol. 71, No. 1, 42-60

SOME METHODOLOGICAL CONSIDERATIONS IN MULTIPLE-CUE PROBABILITY STUDIES[1]

CAROLYN J. HURSCH, KENNETH R. HAMMOND, AND JACK L. HURSCH[2]

University of Colorado

Our principal concern is to make plain that in multiple-cue probability studies carried out within the framework of multiple regression analysis, the statistical properties of the environment and the statistical properties of the response process affect the results. Such statistical properties must be considered when planning multiple-cue probability studies or interpreting their results. Formuli are provided specifying the limitations placed on achievement (inferential accuracy) by statistical factors in certain environments which may be arranged by an E. Illustrations of the application of the analysis from the 2 most representative cases are provided: 1 for probability learning, and 1 for clinical inference.

Within 20 years after Brunswik (1943), Lewin (1943), and Hull (1943) debated the appropriateness of probability laws in psychology, the study of probabilistic behavior has become one of the most vigorous pursuits in psychology. Whatever the upshot of the specific conclusions reached by these three, Brunswik's general argument that probability relationships are not only scientifically respectable, but a useful point of departure for valuable theoretical and empirical work has now been accepted, explicitly or otherwise—especially in the field of learning.

The argument that the environment is an uncertain, probabilistic one, and that inferences about the environment are made on the basis of probabilistic data is fundamental to Brunswik's (1952, 1956) probabilistic functionalism (see also Postman & Tolman, 1959). To use the size constancy paradigm, estimates of object size are based on cues of differing ecological validities, where ecological validity is the correlation between the cue (e.g., linear perspective, interposition, etc.) and the distal variable (size). Cues also differ in their utilization by the subject, where cue utilization is the correlation between cue and inference. Finally, achievement is indicated by the correlation between response and distal variable.

Brunswik used the analogy of a convex lens to describe this situation (see Figure 1). Because both the ecological and the organismic system involve a criterion (distal variable and response, respectively), and because several variables (cues) are correlated in various amounts with both distal variable and response, multiple correlation methods may be applied to the analysis of the interaction of the two systems—and in-

---

[1] This article is based on research at the Behavior Research Laboratory supported by a grant from the National Institute of Mental Health (Project Code Number: M-4977, Clinical Inference and Cognitive Processes) and the council on Research and Creative Activity of the University of Colorado. It is a part of the general program of research on inferential processes now in progress at the Behavior Research Laboratory. Data were analyzed through the facilities of Western Data Processing Center, at University of California, Los Angeles.

[2] The authors are indebted to Frederick J. Todd, Cameron R. Peterson, Daniel E. Bailey, Leon Rappoport, and M. M. Siddiqui for their assistance.

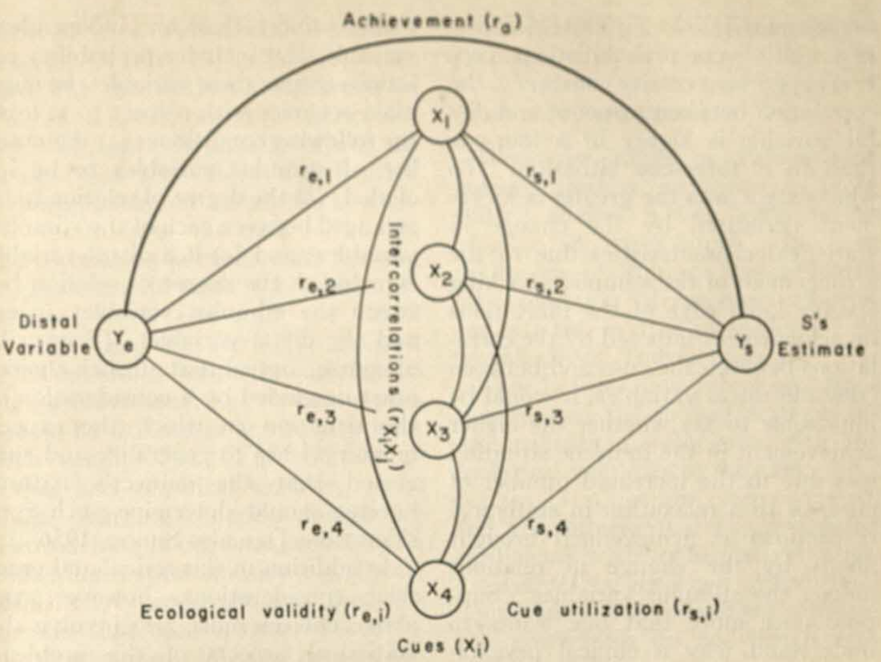


FIG. 1. Brunswik's lens model.

deed this was intended by Brunswik (1956, p. 28; Hammond, 1955; Todd, 1954).

This paper directs itself to the question of *the extent to which the subjects' achievement of a distal variable is determined by the statistical characteristics of the environmental and organismic systems.* Put otherwise, to what extent is inferential accuracy determined by the correlations between cue and distal variable, and between the cues themselves, and to what extent is inferential accuracy determined by the organism's utilization of cues, i.e., the correlations between cue and response. The answers to these questions are here determined within the framework of multiple correlational analysis.

These questions are fundamental to probabilistic theories of behavior which do not restrict themselves to single cue situations, for three reasons: without a knowledge of the limitations placed on achievement by the statistical characteristics of a given ecological or response system, it is impossible (*a*) to *evaluate* a subject's achievement within that system, (*b*) to *compare* a subject's achievement across ecological situations which have different statistical characteristics, and (*c*) to *understand* why the subject's achievement was as high or low as it was.

Suppose, for example, in a multiple-cue probability learning situation the asymptotic level of learning indicates a correlation of .5 between response and distal variable. Lacking information concerning the statistical restrictions placed on the situation by the intercorrelations between the cues, and the correlations between the cues and the distal variable, it would be impossible to say whether the subjects had reached a statistical limit or an organismic limit. Suppose further, in a multiple-cue probabilistic perceptual experiment on size constancy, the correlation between response and distal variable is higher in a four-cue than in a three-cue situation. To what extent was the greater achievement permitted by the change in statistical characteristics due to the arrangement of the stimuli? Lacking specific knowledge of the limitations on achievement imposed by the correlations between the cues and between cues and distal variables, it would be impossible to say whether the higher achievement in the four-cue situation was due to the increased number of cues, or to a relaxation in statistical restrictions on achievement brought about by the change in relations among the stimulus variables. Suppose once more that one wants to understand why a clinical psychologist's inference from test data to patient behavior was not highly accurate. Without specific knowledge of the limitations placed on the clinician's accuracy by his manner of utilizing the cues provided, it would be difficult to suggest ways in which the accuracy of his predictions might be improved.

Knowledge of the above-mentioned statistical characteristics does not, of course, exhaust the prerequisites for understanding learning, perception, or clinical inference, but it is, we believe, one important methodological consideration heretofore given too little attention. The importance of such considerations may be further illustrated by considering the experimenter's choices in arranging stimulus displays, and in evaluating response systems.

Choices in Stimulus Arrangements

Whenever a psychologist wishes to arrange a laboratory experiment which includes more than one independent variable, and includes probability relations among these variables, he must make a choice with respect to at least the following conditions: (*a*) the number of stimulus variables to be included; (*b*) the degree of relation to be arranged between each of the stimulus variables; and (*c*) if a distal variable is included, the degree of relation between the stimulus variables (cues) and the distal variable. (Brunswik, of course, argued that all such choices must be guided by a consideration of the situation to which the experimenter wishes to generalize and suggested that the subject's natural habitat should determine such considerations [see also Simon, 1956].)

In addition to theoretical and practical considerations, however, the above choices must also involve the statistical aspects of the problem. Leaving aside the question of the number of stimulus variables, we shall explore the statistical problems which arise from Choices *b* and *c* above. This implies that multiple-cue probability studies carried out in terms of multiple correlational analysis will require the experimenter to decide on the size of the multiple R controlling the stimulus arrangement. In general terms, the *environmental function-form* should be considered.

Evaluating Response Systems

Assuming that an experimenter has decided upon a certain stimulus arrangement in a multiple-cue probability study, how shall the response system be evaluated? Measurement of achievement can, of course, be made fairly easily by comparing corresponding values of the distal variable and responses made to it—a comparison which can usually be made with a product-moment r (although one must be mindful of Cronbach's, 1955, analysis of interpersonal perception scores). A more difficult problem is: how did the subject's *utilization* of the stimulus information affect his achievement? Brunswik (1956) assumed that the subject's utilization coefficients should match the ecological validities of the cues. That is, the correlation between each cue and response should equal the correlation between each cue and the distal variable, and he certainly assumed that matching would be related to high achievement, e.g., "Ideally, cues should be utilized in accordance with their validity [p. 141]." The theoretical and empirical work by probability learning theorists concerning the matching of response probabilities with stimulus probabilities (e.g., Estes, 1959) leaves no doubt but that the question has been considered worthy of investigation. We shall explore the statistical aspects of matching behavior within the multiple correlation framework. Within the idiom of probabilistic functionalism, we shall explore the statistical aspects of the match between environmental *function-form* and the *process-form* of the subject's response system.

Linear versus Nonlinear Relationships

There is one further aspect of multiple-cue probability studies to be considered—the linearity-nonlinearity question, i.e., the component-pattern issue. Experimenters must decide to what extent the relations among the cues should be linear or patterned, and they must be prepared to examine the extent to which the subject's behavior can be explained by a linear or nonlinear model (Estes, 1959; Hammond, 1955; Hoffman, 1960; Meehl, 1960; Sarbin, Taft, & Bailey, 1960; and others). This matter will also be considered, and relevant data from clinical psychologists will be examined.

Cases to be Considered

We have taken a Brunswikian posture toward multiple-cue probability studies which leads us to examine six cases which arise from (*a*) choices in stimulus arrangements, (*b*) evaluation of response systems, and (*c*) the matter of linearity and nonlinearity. The six cases are:

Case 1: Cues have intercorrelations of zero, and the subject matches his cue utilization coefficients with the ecological validities of the cues.

Case II: Cues are intercorrelated, and the subject matches utilization coefficients with ecological validities.

Case III: Cues have intercorrelations of zero, and the subject's utilization coefficients do not match ecological validities.

Case IV: Cues are intercorrelated, and the subject's utilization coefficients do not match ecological validities.

Case V: Cues have intercorrelations of zero, and the differences between the subject's utilization coefficients and ecological validities are evaluated.

Case VI: Same as Case V except that cues are intercorrelated.

Furthermore, in the development of the six cases described above, we shall consider two subcases: (*a*) one in

TABLE 1

NATURE OF PRESENTATION OF THE ANALYSIS OF SIX RELATED CASES

Relation between ecological validities ($r_{e,i}$) and utilization coefficients ($r_{s,i}$)	Relations between cues (rx_ix_j) Intercorrelations = 0 ($rx_ix_j = 0$)	Intercorrelations ≠ 0 ($rx_ix_j \neq 0$)
Matched	Case I	Case II
Not matched	Case III	Case IV
Degree of matching	Case V	Case VI

which the situation is described by linear relationships among all the variables, and (*b*) one in which nonlinear relations must be investigated. The nature of the presentation is indicated in Table 1.

Definitions

Assume that

$$Y_e, Y_s, X_1, \cdots, X_n \qquad [1]$$

are standardized random variables (i.e., their mean is zero and their variance is 1) where (see Figure 1):

Y_e is the distal variable in the environment

Y_s is the subject's estimate of the distal variable

$X_1, \cdots, X_n$ are the cue variables

This assumption does not decrease the applicability of the results since any random variable x may be "standardized" by replacing it with the random variable $X = x - m/s$ where m stands for the mean of x and s^2 is the variance of x.

$$\begin{aligned} \text{Correlation } (Y_e, X_i) &= r_{e,i} \\ \text{Correlation } (Y_s, X_i) &= r_{s,i} \end{aligned} \qquad [2]$$

Let R_e be the multiple correlation of Y_e and $(X_1, \cdots, X_n)$ on the environment side, [3]

Let R_s be the multiple correlation of Y_s and $(X_1, \cdots, X_n)$ on the subject's side.

Under these assumptions, Y_e and Y_s have the following representation:

$$\begin{aligned} Y_e &= (\beta_{e,1}X_1 + \cdots + \beta_{e,n}X_n) + Z_e \\ Y_s &= (\beta_{s,1}X_1 + \cdots + \beta_{s,n}X_n) + Z_s \end{aligned} \qquad [4]$$

where the quantities in parentheses represent the regressions, and Z_e and Z_s are the residuals. Z_e and Z_s are uncorrelated with any of the regression variables $(X_1, \cdots, X_n)$.

Then

$$\begin{aligned} R_e^2 &= \beta_{e,1}r_{e,1} + \cdots + \beta_{e,n}r_{e,n} \\ &= \sum \beta_{e,i}r_{e,i} \\ R_s^2 &= \beta_{s,1}r_{s,1} + \cdots + \beta_{s,n}r_{s,n} \\ &= \sum \beta_{s,i}r_{s,i} \end{aligned} \qquad [5]$$

Also, $EZ_e = 0$, $EZ_s = 0$ (where E indicates "expected value")

$$\begin{aligned} \text{var } Z_e &= 1 - R_e^2, \\ \text{var } Z_s &= 1 - R_s^2 \end{aligned} \qquad [6]$$

$$r_a = E(Y_eY_s) = \text{Cov } (Y_eY_s), \qquad [7]$$

and is a correlation coefficient represented by the lens model in Figure 1, as "achievement."

Let the correlation $(Z_e, Z_s) = C$. The C is the partial correlation between Y_e and Y_s when the effects of $X_1, \cdots, X_n$ are eliminated, and may be described as the correlation between the variance unaccounted for by the multiple correlation in the ecology and the variance unaccounted for by the multiple correlation in the subject's response system.

The assumption is often made that residual variance is random, and, therefore, that the correlation between two sets of residual variances must be zero. This assumption is not a necessary feature of regression analysis (Siddiqui, 1960). The residual variances may be due to random error *or* to nonlinearity (pattern) in the system. Therefore, randomness is but one of the many possible reasons for the existence of residual variance; to assume that randomness is the only reason overlooks the patterning or nonlinearity that might account for residual variance. (This point is discussed below in more detail.)

The variable C is important because achievement may be a result of the subject's perception of a nonlinear relation between cues and distal variable. If such a nonlinear relation developed between cues and response, it would not, of course, be reflected in the subject's multiple correlation coefficient. Although the value of C would not define the subject's specific nonlinear *method* of combining data, it would provide the following information:

1. If C were high and positive, this would indicate that the subject is correctly utilizing a nonlinear relationship between the cues and the distal variable. "Correctly" here means that the subject's dependency on the cues follows the same nonlinear form that is present in the cue-to-distal-variable relationship. This would contribute to high achievement.
2. If C were high and negative, this would indicate that the subject is using a negative function of the nonlinear relationship between cues and distal object. This would contribute to low achievement.
3. If C were zero, this would indicate that the subject's residual variance is derived from the use of a nonlinear function which is not the same as that in the environment, or that one or both of the residual variances are random.
4. Naturalistic studies provide a special case, for cues may be present in the environment and the subject may be utilizing them although the experimenter may not be aware of their existence. Thus, in naturalistic experiments where the subject can perceive cues other than those denoted by the experimenter, the subject may be utilizing a cue which has a high linear correlation with the distal variable. A high value of C would therefore be obtained, and would be misinterpreted. This point has been discussed by Hoffman (1960). In a contrived stimulus arrangement, however, the intrusion of valid cues unknown to the experimenter is not likely.

In all cases, the usefulness and significance of the variable C will decrease as the value of the environmental R approaches 1.00.

The limits of accuracy (r_a) will now be derived for the six cases. Each case will be developed first under the condition that $C = 0$, and then under the condition that $C \neq 0$.

Case I

Conditions:

(*a*) $r_{X_iX_j} = 0, \quad i \neq j$, and

(*b*) $r_{e,i} = r_{s,i}, \quad$ for all i.

The reader may be reminded that the relation between correlation coefficients and beta weights is as follows:

$$\beta_{e,1} = \frac{r_{e1} - r_{e2}r_{12}}{1 - r_{12}^2}$$

where $r_{e,1}$ is the correlation between the distal variable Y and the cue X_1

$r_{e,2}$ is the correlation between the distal variable Y and the cue X_2

$r_{1,2}$ is the intercorrelation between the cues X_1 and X_2.

This expression is, of course, more complicated when there are more than $2X_i$. For $n > 2$, the reader should consult McNemar (1962). Also note that when the X_i are independent, the intercorrelations are zero, and $\beta_{e,i} = r_{e,i}$.

When the X_i are independent, and matching occurs, i.e., if the subject's utilization coefficients are the same as the ecological validities, then

$$R_e^2 = R_s^2 = R^2, \qquad \text{and} \qquad r_{e,i} = r_{s,i} = r_i$$

and

$$r_a = E(Y_e Y_s)$$

$$E(Y_e Y_s) = E[(\beta_{e,1}X_1 + \cdots + \beta_{e,n}X_n + Z_e)(\beta_{s,1}X_1 + \cdots + \beta_{s,n}X_n + Z_s)]$$

$$= E[(\sum_{i,j=1,n} \beta_{e,i}\beta_{s,j}X_iX_j) + E(Z_eZ_s)]$$

$$= \sum_{i,j=1,n} \beta_{e,i}\beta_{s,j}E(X_iX_j) + E(Z_eZ_s)$$

since $E(X_iX_j)$ in this case $= 0$ for $i \neq j$

and

$$E(X_iX_i) = E(X_i^2) = 1$$

then

$$E(Y_eY_s) = \sum \beta_{e,i}\beta_{s,i} + E(Z_eZ_s)$$

but since in this case

$$\beta_{e,i} = r_i = \beta_{s,i}$$

we have

$$r_a = r_1^2 + \cdots + r_n^2 + E(Z_eZ_s) \qquad [8]$$

where

$$E(Z_eZ_s) = C\sigma_{Z_e}\sigma_{Z_s}$$

and

$$\sigma_{Z_e} = \sqrt{1 - R_e^2}; \qquad \sigma_{Z_s} = \sqrt{1 - R_s^2}$$

1. If $C = 0$,

then

$$r_a = r_1^2 + r_2^2 + \cdots + r_n^2 = R^2. \qquad [9]$$

In this case, as in an orthogonal component, factorial design, the cues are uncorrelated. Also, the subject matches utilization coefficients with validity coefficients. His multiple correlation coefficient, therefore, is equal to that in the environment. The limits of achievement are determined by the size of the (equal) multiple correlation coefficients (R_e^2 and R_s^2) and the correlation (C) between the variance unaccounted for by R_e^2 and the variance unaccounted for by R_s^2. If the correlation between these variances is 0 (as in Formula 9) then achievement is completely determined by the size of the squared multiple correlation coefficients. In short, should a subject adapt to a completely linear environment by matching his utilization coefficients with the ecological validities of the cues, his achievement will be precisely the square of the multiple correlation of the environment. Although the result may be perfectly obvious algebraically, it has not been obvious in experimental settings. Variation in environmental R, therefore, should become an important experimental variable in multiple-cue probability learning and perceptual experiments. Without such variation, results and conclusions remain highly situation-bound.

2. If $C \neq 0$,

then from Formula 8

$$r_a = R^2 + C(1 - R^2). \qquad [10]$$

If C is positive, then, achievement (r_a) will be increased as a function of the value of C. That is, to the extent that a pattern, or nonlinear arrangement exists among the cues in the environment, and to the extent that S utilizes this pattern correctly, his achievement will increase. If C is negative, r_a, of course, will be reduced and may even become negative. Formula 10 provides the mathematical explication of our earlier remarks about the value of investigating C, the nonlinear aspect of the achievement variable. It also illustrates the fact that it is possible to partition the component and pattern determinants of the subject's response. Since $-1 \leq C \leq 1$, the limits of achievement in this case are given by

$$2R^2 - 1 \leq r_a \leq 1. \quad [11]$$

Case II

Conditions:

(a) $r_{X_iX_j} \neq 0$; $i \neq j$;

(b) $r_{e,i} = r_{s,i}$, for all i.

When the X_i are not independent, and the utilization coefficients match the ecological validity coefficients, then

$$r_{e,i} = r_{s,i} = r_i \text{ for all } i$$

and

$$\beta_{e,i} = \beta_{s,i} \text{ for all } i,$$

because the β_i are completely determined by the r_i and the $r_{X_iX_j}$. (See also Footnote 3.)

Thus

$$R_e^2 = R_s^2 = R^2$$

also,

$$R^2 = \beta_1 r_1 + \cdots + \beta_n r_n = \sum \beta_i r_i$$

and

$$r_a = \sum \beta_i r_i + C(1 - R^2). \quad [12]$$

1. If $C = 0$,

then

$$r_a = \sum \beta_i r_i = R^2. \quad [13]$$

This case removes the restriction that the cues must be uncorrelated with one another. Matching still prevails, however, which leaves us with the same result as in the first subcase of Case I. Leaving cue variables "tied" in their natural ecological form, then, has no effect on the limits of achievement (providing matching occurs and $C = 0$). We now consider the subcase where $C \neq 0$.

2. If $C \neq 0$, then because $-1 \leq C \leq 1$, from Formula 12, the limits of r_a are:

$$\sum \beta_i r_i - (1 - R^2) \leq r_a \leq \sum \beta_i r_i + (1 - R^2). \quad [14]$$

Since $\sum \beta_i r_i = R^2$, the value of C (i.e., the degree of correlation between the nonlinear aspect of the environment and the nonlinear aspect of the subject's response process) will add to or subtract from achievement due to linearity. More specifically, if there is a pattern in the environment which is correctly used by the subject, then achievement will be increased precisely by the quantity $C(1 - R^2)$.

---

[3] The beta weight enters here for the first time as a quantity not equal to r in the computation of each R and therefore each r_a. This is due to Condition (a) above, which indicates that the cues (the X_i) are intercorrelated, not uncorrelated as they were in Case I. The beta weight for the ecological R combines an ecological validity coefficient ($r_{e,i}$) with the intercorrelations of each cue with the i^{th} cue. The beta weight for the subject's response process-form combines a utilization coefficient ($r_{s,i}$) and the intercorrelations of each cue with the i^{th} cue. When the intercorrelations are all zero, then the beta weight β_i equals the corresponding validity coefficient r_i, and $\beta_i r_i = (r_i)^2$.

Case III

Conditions:

(*a*) $r_{X_iX_j} = 0, \quad i \neq j$, and

(*b*) $r_{e,i} \neq r_{s,i}$, for all i.

When the X_i are independent, then the $\beta_{e,i}$ and $\beta_{s,i}$ in Formula 5 reduce to $r_{e,i}$ and $r_{s,i}$, respectively, and Formula 5 becomes

$$R_e^2 = \sum r_{e,i}^2; \quad R_s^2 = \sum r_{s,i}^2; \qquad [5a]$$

then,

$$r_a = E(Y_eY_s)$$

$$\begin{aligned} r_a = r_{e,1}r_{s,1} & \\ & + r_{e,2}r_{s,2} + \cdots + r_{e,n}r_{s,n} \\ & + C\sqrt{(1 - R_e^2)(1 - R_s^2)} \qquad [15] \end{aligned}$$

1. If $C = 0$, then from Schwarz's inequality[4]

$$\begin{aligned} |r_{e,1}r_{s,1} + \cdots + r_{e,n}r_{s,n}| & \\ & \leq \sqrt{\sum r_{e,i}^2 \sum r_{s,i}^2} = R_eR_s \qquad [16] \end{aligned}$$

This gives the following bounds on r_a:

$$-R_eR_s \leq r_a \leq +R_eR_s \qquad [17]$$

We conclude then, that under these circumstances, achievement is bounded by the *product* of the two multiple correlations involved. Note that should either multiple R be 1.00, the bounds on achievement are plus and minus the value of the other R. This observation is important because investigators frequently present the subject with an ecology which is completely determined, i.e., $R_e = 1.00$. (An example will be presented under Case V.) Thus Brunswik's emphasis on the importance of environmental stipulations is borne out.

Also, note that if there is no possibility for a positive C in a given environment, a subject may increase his achievement by maintaining a response system which is more linear than the environment. In other words, under the conditions of this Case, if R_e is less than 1.00, then matching cue utilization coefficients with validity coefficients will give as an upper limit, $r_a = R^2$, while a completely linear response system, ($R_s^2 = 1.00$) will give the subject an upper limit of $r_a = R_e$, which, of course, is higher than R_e^2 when R_e^2 is less than 1.00. However, see Cases V and VI for a more detailed treatment of this point.

Formula 17 also illustrates the importance of the experimental variation of the multiple correlation in the environment. Suppose that the linear multiple correlation model does indeed account for all of the variation in a subject's inferential behavior; then the upper bound on his maximum achievement is reduced to the extent that the environmental R departs from unity. Should one experimenter arrange his stimulus display so that the environmental $R = .7$, and a second experimenter set his environmental $R = .3$, then the upper bound on the subjects' potential achievements would be different in the two experiments, even though the multiple regression equation provided an adequate account of all subjects' inferential behavior in both situations. In short, adequate environmental control in multiple-cue probability studies requires the specification and manipulation of the environmental multiple R. The authors know of few such considerations in the literature. (see, however, Azuma & Cronbach, 1961; Lee & Tucker, 1962; Rappoport, 1961; Smedslund, 1955; Summers, 1962; Todd, 1954.)

2. If $C \neq 0$, then since $-1 \leq C \leq 1$, the limits of r_a (using Formula 15)

---

[4] The application of Schwarz's inequality to this problem is shown in the Appendix to this article.

are:

$$\sum r_{e,i}r_{s,i} - \sqrt{(1 - R_e^2)(1 - R_s^2)} \leq r_a \leq \sum r_{e,i}r_{s,i} + \sqrt{(1 - R_e^2)(1 - R_s^2)}. \quad [18]$$

According to Formula 18, if the residual variances Z_eZ_s are correlated, the limits of achievement are determined by the values of the ecological validities ($r_{e,i}$), the utilization coefficients ($r_{s,i}$), the multiple *R*s (R_e and R_s), and *C*. Most important, the complexity of Formula 18 means that when $C \neq 0$, the limits placed on the subjects' achievement by the investigator will not be apparent to him until he has made the calculations indicated. Therefore, they must be made in order to evaluate the subject's achievement.

Case IV

Conditions:

(*a*) $r_{x_i x_j} \neq 0, \quad i \neq j$;

(*b*) $r_{e,i} \neq r_{s,i}$ for all *i*.

When the X_i are not independent, and the utilization coefficients do not match the ecological validity coefficients, then

$$R_e^2 = \beta_{e,1}r_{e,1} + \cdots + \beta_{e,n}r_{e,n}$$

and

$$R_s^2 = \beta_{s,1}r_{s,1} + \cdots + \beta_{s,n}r_{s,n}. \quad [5]$$

Since, from Formula 4,

$$Y_e = \sum \beta_{e,i}X_i + Z_e,$$

and

$$Y_s = \sum \beta_{s,i}X_i + Z_s,$$

and Y_e and Y_s are normalized, therefore $r_a = E(Y_e, Y_s)$

$$r_a = \sum_i \sum_j \beta_{s,i}\beta_{e,j}E(X_i, X_j) + E(Z_e, Z_s)$$

$$= \sum_i \beta_{s,i}\Big(\sum_j \beta_{e,j}r_{x_i x_j}\Big) + r_{z_e z_s}\sigma_{z_e}\sigma_{z_s}$$

and since

$$\Big(\sum_j \beta_{e,j}r_{x_i x_j}\Big) = \beta_{e,1}r_{x_i x_1} + \beta_{e,2}r_{x_i x_2} + \cdots + \beta_{e,n}r_{x_i x_n} = r_{e,i};$$

then,

$$r_a = \sum \beta_{s,i}r_{e,i} + C\sqrt{(1 - R_e^2)(1 - R_s^2)} \quad [19]$$

1. If $C = 0$,

then

$$r_a = \sum \beta_{s,i}r_{e,i} = \sum \beta_{e,i}r_{s,i}. \quad [20]$$

Achievement in this case is determined by the ecological validity coefficients and the intercorrelations between the cues which may be arranged by the *experimenter*, and by the correlations between cue and inference which are determined by the *subject*. Obviously then, the experimenter, by his choice of validity coefficients and intercorrelations, can greatly affect the subject's achievement. For achievement depends upon a combination of the environmental validity coefficients, and the intercorrelations between cues as well as the subject's cue utilization coefficients, all of which are contained in the expressions $\sum \beta_{s,i}r_{e,i}$ and $\sum \beta_{e,i}r_{s,i}$.

2. If $C \neq 0$, then as $-1 \leq C \leq 1$, the limits of r_a are:

$$\sum \beta_{s,i}r_{e,i} - \sqrt{(1 - R_e^2)(1 - R_s^2)} \leq r_a \leq \sum \beta_{s,i}r_{e,i} + \sqrt{(1 - R_e^2)(1 - R_s^2)}. \quad [21]$$

Thus, if an experimenter happens to have arranged the ecology in such a way that achievement will be low if the subject considers only linear relationships between cues and the distal variable, then high achievement will be due to a high correlation between the residual pattern variance in the environment and the residual pattern variance in the subject's response.

The first four cases present the logic for determining the limits of accuracy when a subject either matches or does not match utilization coefficients with validity coefficients, and the formuli and rationale necessary to the solutions of the more general situations to be considered in Cases V and VI. Since perfect matching would probably occur very infrequently in any empirical situation, and nonmatching could occur to any extent, the more realistic approach of Cases V and VI involving the *degree* of matching is considered.

Case V

Conditions:

(a) $r_{x_ix_j} = 0;\quad i \neq j;$

(b) $\sum (r_{e,i} - r_{s,i})^2 = d_r^2.$

As in Formula 5a, when $r_{x_ix_j} = 0$, then

$$R_e^2 = \sum (r_{e,i})^2,$$

and

$$R_s^2 = \sum (r_{s,i})^2.$$

If complete matching occurs, i.e., if $r_{e,i} = r_{s,i}$ for all i, then $d_r^2 = 0$, as in Case I.

However, if all of the utilization coefficients do not match the validity coefficients, then

$$\begin{aligned} d_r^2 &= \sum (r_{e,i} - r_{s,i})^2 \\ &= \sum (r_{e,i})^2 - 2\sum (r_{e,i}r_{s,i}) \\ &\quad + \sum (r_{s,i})^2 \\ &= R_e^2 - 2\sum r_{e,i}r_{s,i} + R_s^2 \end{aligned}$$

or

$$\sum (r_{e,i}r_{s,i}) = \frac{R_e^2 + R_s^2 - d_r^2}{2} \qquad [22]$$

From Case III, Formula 15 we know that

$$r_a = \sum r_{e,i}r_{s,i} + C\sqrt{(1-R_e^2)(1-R_s^2)}$$

therefore

$$r_a = \frac{R_e^2 + R_s^2 - d_r^2}{2} + C\sqrt{(1 - R_e^2)(1 - R_s^2)}. \qquad [23]$$

If $d_r^2 = 0$ (as in Case I) then

$$R_e^2 = R_s^2 = R^2$$

and

$$r_a = R^2 + C(1 - R^2). \qquad [24]$$

1. (*a*) If $C = 0$, and $d_r^2 = 0$, then (as in Case I) $r_a = R^2$.

   (*b*) If $C = 0$, and $d_r^2 \neq 0$, then
   $$r_a = \frac{R_e^2 + R_s^2 - d_r^2}{2}. \qquad [25]$$

2. (*a*) If $C \neq 0$, and $d_r^2 = 0$, then (as in Case I, Formula 11) the limits of r_a are:
   $$2R^2 - 1 \leq r_a \leq 1.$$

   (*b*) If $C \neq 0$, and $d_r^2 \neq 0$, then the limits of r_a are:

$$\frac{R_e^2+R_s^2-d_r^2}{2} - \sqrt{(1-R_e^2)(1-R_s^2)} \leq r_a \leq \frac{R_e^2+R_s^2-d_r^2}{2} + \sqrt{(1-R_e^2)(1-R_s^2)} \qquad [26]$$

In this case, the cues are uncorrelated and achievement is shown to be a function of the squared differences[5] between the utilization coefficients and the validity coefficients, as well as R_e^2 and R_s^2. As this difference approaches zero, achievement will approach $\frac{R^2_e + R^2_s}{2}$, and at the same time be dependent upon the degree of correlation of the two residuals. (Note again that if $R_e = 1.00$, then r_a is completely determined by

---

[5] The quantity d_r^2 used in Case V to indicate differences between utilization coefficients and validity coefficients is equal to the quantity d_rd_β used in Case VI to indicate such differences. Case V deals with uncorrelated cues ($r_{x_ix_j} = 0$), and any β_i is equal to the corresponding r_i under this condition. Therefore, given that $r_{x_ix_j} = 0$, then $\beta_i = r_i$, and $d_r^2 = d_rd_\beta$.

$$\left.\frac{R^2_e + R^2_s - d_r^2}{2}.\right)$$

We turn now to an illustration of Case V in which the environmental $R = 1.00$, intercorrelations between cues are zero, and where the *degree* to which the subject's utilization coefficients match the ecological validity coefficients of the cues is considered. Summers' (1962) multiple probability learning experiment provides appropriate data.

In Summers' experiment there were 3 cues (color, angle, and area) whose correlations with the distal variable (ecological validities) were .743, .371, and .557. Six different conditions provided six combinations of these three values with the three cues; e.g., in Condition I, the correlations of cue with distal object were: color .743; angle .557; and area .371; the next condition interchanged the correlations for angle and area, and so forth. Different subjects were used for each condition. There were six blocks of trials in each condition, the last block consisting of an extinction series. Thus, Summers' data contain the necessary conditions for a simple application of the formuli of Case V, where $r_{X_iX_j} = 0$, and $\sum (r_{e,i} - r_{s,i})^2 = d_r^2$, and where $d_r \neq 0$. The ecological multiple correlation $(R_e) = 1.00$. (The example following Case VI illustrates the situation where $R_e \neq 1.00$.)

There are two points to be observed from the application of the formuli of Case V to Summers' data: (*a*) r_a, the correlation between the subject's estimate of the distal variable and the actual value of it, may be computed from Formula 25 rather than from the correlation coefficient between the judgment and distal variable from the raw data as Summers did. If the logic of the derivation of Formula 25 is correct, r_a computed from Formula 25 should be exactly the same as r_a computed by Summers. It is, for all of the 17 subjects whose data we analyzed. (*b*) The application of Formula 25 should throw some light on the nature of a subject's performance. We will now demonstrate that it does. Formula 25 reads, it will be recalled, as follows:

$$r_a = \frac{R_e^2 + R_s^2 - d_r^2}{2}.$$

TABLE 2

APPLICATION OF EQUATION 25 TO SUBJECT NO. 6 IN SUMMERS' EXPERIMENT

$R_e^2 = 1.00$

$R_s^2 = .970$ for Block I $\quad d_r^2 = 3.002$

$R_s^2 = .203$ for Block V $\quad d_r^2 = .634$

$R_s^2 = .212$ for Block VI $\quad d_r^2 = .616$

Block I $\quad r_a = \dfrac{1.00 + .970 - 3.002}{2} = -.516$

Block V $\quad r_a = \dfrac{1.00 + .203 - .634}{2} = .284$

Block VI $\quad r_a = \dfrac{1.00 + .212 - .616}{2} = .298$

The relevant data for Subject No. 6 are presented in Table 2.

Note that during the first block of trials, the manner in which Subject No. 6 combined the data from the stimulus array essentially could be represented by the multiple correlation procedure; the multiple correlation (R_s^2) between the cues and his response being .970. Thus, Subject No. 6 functions in almost the identical manner in which the environment, about which he must learn, functions. But note also that although Subject No. 6's response process-form closely matches the form in which the environment functions, it is badly

TABLE 3

APPLICATION OF EQUATION 25 TO SUBJECT NO. 12 IN SUMMERS' EXPERIMENT

$R_e^2 = 1.00$

$R_s^2 =$ .523 for Block I $\quad d_r^2 = 2.092$

$R_s^2 =$ .306 for Block V $\quad d_r^2 = .456$

$R_s^2 =$ .310 for Block VI $\quad d_r^2 = .379$

$$\text{Block I} \quad r_a = \frac{1.00 + .523 - 2.092}{2} = -.284$$

$$\text{Block V} \quad r_a = \frac{1.00 + .306 - .456}{2} = .425$$

$$\text{Block VI} \quad r_a = \frac{1.00 + .310 - .379}{2} = .466$$

applied—the difference between ecological validities and utilization coefficients (d_r^2) being very large (3.002). As a result, achievement in the first block of trials is poor; r_a is a fairly large *negative* coefficient, −.516. However, by the time Block V is reached, Subject No. 6 has largely given up a linear, additive procedure ($R_s^2 = .203$ instead of .970); also d_r^2 has been markedly reduced (from 3.002 to .634), and as a result achievement has increased (r_a now equals +.284 instead of −.516). No marked change took place in Block VI. In short, an increase in achievement came about as a result of *decreasing* the match between the subject's response process-form (R_s^2) and the environmental function-form (R_e^2), and *increasing* the match between utilization coefficients and ecological validities.

Subject No. 12, (Table 3), behaves in much the same way as Subject No. 6 behaved in Block I and Block V. That is, Subject No. 12 increases his achievement despite *decreasing* the match between the environmental function-form (R_e^2) and his response process-form (R_s^2), for he more than makes up for this by *increasing* the match between utilization coefficients and ecological validities (reducing d_r^2). In Block VI, Subject No. 12 maintains the former match (.310 being virtually identical with .306) but by further decreasing d_r^2, slightly increases r_a from .425 to .466.

Subject No. 7 provides an illustration of a different kind of performance by (see Table 4) decreasing a small d_r^2 (from .962 to .019), and increasing his multiple R^2 from essentially zero (.003) to a rather high R^2 (.783) which increased r_a from near zero (.020) to +.882. Thus, Subject No. 7 increased his achievement by increasing the match between his response process-form and the environmental function-form, as well as increasing the match between ecological validities and utilization coefficients.

In short, our analysis may be used to throw light on a subject's performance. In the above examples, we could see that subjects can increase their achievement by different methods, either by increasing the match between environmental function-form and response process-form, and/or by increasing the match between utilization coefficients and ecological validities. Differing possibilities such as these seem simple, and perhaps obvious. At the same time, they raise fundamental questions: Can individuals *learn* to match response process-

TABLE 4

APPLICATION OF EQUATION 25 TO SUBJECT NO. 7 OF SUMMERS' EXPERIMENT

$R_e^2 = 1.00$

$R_s^2 =$ .003 for Block I $\quad d_r^2 = .962$

$R_s^2 =$ .783 for Block VI $\quad d_r^2 = .019$

$$\text{Block I} \quad r_a = \frac{1.00 + .003 - .962}{2} = .020$$

$$\text{Block VI} \quad r_a = \frac{1.00 + .783 - .019}{2} = .882$$

forms with environmental function-forms? This seems to be analogous to the learning of principles. It is certainly what Brunswik would have called "distal focusing." The above analysis shows that not only can function-forms be matched, but that they can be misapplied (Subject No. 6, Block I). Can individuals learn to match ecological validities with utilization coefficients?[6] This question seems to concern proximal rather than distal focusing, but its importance remains. Whatever the case, when examining the performance of an individual in a multiple-cue probability situation, it will be necessary to consider *both* the match between response process-form and environmental function-form and the match between ecological validities and cue utilization coefficients. Formula 25 makes this possible.

Case VI

Conditions:

(a) $r_{x_i x_j} \neq 0, \quad i \neq j;$

(b) $d_\beta = \sqrt{\sum (\beta_{e,i} - \beta_{s,i})^2};$

(c) $d_r = \sqrt{\sum (r_{e,i} - r_{s,i})^2}.$

If complete matching of utilization coefficients with validity coefficients occurs, then

$$d_r = 0, \quad \text{and} \quad d_\beta = 0.$$

If utilization coefficients do not match validity coefficients, then

$$r_a = \frac{R_e^2 + R_s^2 - (\beta_{e,i} - \beta_{s,i}, r_{e,i} - r_{s,i})}{2} + C\sqrt{(1 - R_e^2)(1 - R_s^2)} \quad [27]$$

or, as C runs from -1 to 1,

$$\frac{R_e^2 + R_s^2 - (\beta_{e,i} - \beta_{s,i}, r_{e,i} - r_{s,i})}{2} - \sqrt{(1 - R_e^2)(1 - R_s^2)} \leq r_a$$
$$\leq \frac{R_e^2 + R_s^2 - (\beta_{e,i} - \beta_{s,i}, r_{e,i} - r_{s,i})}{2} + \sqrt{(1 - R_e^2)(1 - R_s^2)} \quad [27a]$$

where

$$(\beta_{e,i} - \beta_{s,i}, r_{e,i} - r_{s,i}) = \sum (\beta_{e,i} - \beta_{s,i})(r_{e,i} - r_{s,i})$$

however, by Schwarz's inequality

$$|(\beta_{e,i} - \beta_{s,i}, r_{e,i} - r_{s,i})| \leq d_\beta d_r$$

and the bounds for r_a are:

$$\frac{R_e^2 + R_s^2 - d_\beta d_r}{2} + C\sqrt{(1 - R_e^2)(1 - R_s^2)} \leq r_a$$
$$\leq \frac{R_e^2 + R_s^2 + d_\beta d_r}{2} + C\sqrt{(1 - R_e^2)(1 - R_s^2)}.$$

If

$$(\beta_{e,i} - \beta_{s,i}, r_{e,i} - r_{s,i}) = 0 \quad \text{then} \quad R_e^2 = R_s^2 = R^2$$

and

$$r_a = R^2 + C(1 - R^2).$$

---

[6] Recent work by Newton, (1964) at the Behavior Research Laboratory suggests that subjects can.

1. (*a*) If $C = 0$, and $d_r = 0$

then

$$r_a = R^2.$$

(*b*) If $C = 0$, and $d_r \neq 0$

then

$$r_a = \frac{R_e^2 + R_s^2 - (\beta_{e,i} - \beta_{s,i}, r_{e,i} - r_{s,i})}{2}$$

and the bounds for r_a are:

$$\frac{R_e^2 + R_s^2 - d_\beta d_r}{2} \leq r_a \leq \frac{R_e^2 + R_s^2 + d_\beta d_r}{2}. \qquad [28]$$

2. (*a*) If $C \neq 0$, and $d_r = 0$

then the limits of r_a are (as in Case I, Formula 11):

$$2R^2 - 1 \leq r_a \leq 1.$$

(*b*) If $C \neq 0$, and $d_r \neq 0$, then since $-1 \leq C \leq 1$ the bounds of r_a are:

$$\frac{R_e^2 + R_s^2 - d_\beta d_r}{2} - \sqrt{(1 - R_e^2)(1 - R_s^2)} \leq r_a$$
$$\leq \frac{R_e^2 + R_s^2 + d_\beta d_r}{2} + \sqrt{(1 - R_e^2)(1 - R_s^2)}. \qquad [29]$$

This case is the same as Case V except that the cues are correlated; therefore, the beta weight is introduced into the consideration of differences between the subject's multiple correlation and the environmental multiple correlation. The difficulty of obtaining the variable C by empirical methods may be overcome by manipulation of the algebraic relationships already presented, as follows:

Since (from Formula 27)

$$r_a = \frac{R_e^2 + R_s^2 - \sum (\beta_{e,i} - \beta_{s,i})(r_{e,i} - r_{s,i})}{2} + C\sqrt{(1 - R_e^2)(1 - R_s^2)}$$

then

$$C = \left(r_a - \frac{R_e^2 + R_s^2}{2} + \frac{\sum (\beta_{e,i} - \beta_{s,i})(r_{e,i} - r_{s,i})}{2}\right) \frac{1}{\sqrt{(1 - R_e^2)(1 - R_s^2)}}. \qquad [30]$$

This formula may be used for computations.

As an illustration of the utility of the analysis presented under Case VI, data from a study by Todd (1954; see also Hammond, 1955) have been analyzed. Ten clinicians made judgments of the IQ scores of 78 patients after an examination of the patients' scores on 19 Rorschach factors. In order to measure achievement, Wechsler-Bellevue IQ scores were obtained for the 78 patients. Multiple regression equations were obtained for the environment by relating the Wechsler-Bellevue scores (distal variable) to the various scores on the Rorschach factors (cues). Multiple regression

TABLE 5

DATA FROM TODD (1954) SHOWING THE EFFECT OF ENVIRONMENTAL VARIABLES ON ACHIEVEMENT

Clinician	R_s^2	$\Sigma(\beta_{e,i}-\beta_{s,i})(r_{e,i}-r_{s,i})$	C*	Limits of r_a	r_a	r_a if $C=0$
1	.8585	.774	−.12	.10 to .58	.316	.345
2	.7126	.723	−.12	−.04 to .64	.255	.295
3	.8892	.725	−.05	.18 to .58	.374	.385
4	.8924	.793	.14	.15 to .55	.378	.350
5	.7779	.984	−.20	−.10 to .50	.142	.200
6	.9464	.795	.30	.23 to .52	.419	.375
7	.8271	1.090	−.19	−.09 to .43	.121	.170
8	.9038	1.041	−.43	.03 to .43	.144	.230
9	.8403	.838	−.38	.05 to .55	.204	.300
10	.7666	.540	.17	.11 to .71	.465	.414
11[a]	.6022	0	—	.20 to 1.00	—	.602

Note.—Computational example:

Clinician No. 1

$$R_e^2 = .6022$$
$$R_s^2 = .8585$$
$$r_a = .316$$
$$\Sigma(\beta_{e,i}-\beta_{s,i})(r_{e,i}-r_{s,i}) = .7740$$

from Formula 27a

$$\frac{R_e^2 + R_s^2 - \Sigma(\beta_{e,i}-\beta_{s,i})(r_{e,i}-r_{s,i})}{2} - C\sqrt{(1-R_e^2)(1-R_s^2)} \le r_a \le \frac{R_e^2 + R_s^2 - \Sigma(\beta_{e,i}-\beta_{s,i})(r_{e,i}-r_{s,i})}{2} + C\sqrt{(1-R_e^2)(1-R_s^2)}.$$

Since the limits of C are: $-1 \le C \le 1$, the limits of r_a for Clinician No. 1 are:

$$\frac{.602 + .858 - .774}{2} - \sqrt{(.398)(.142)} \le r_a \le \frac{.602 + .858 - .774}{2} + \sqrt{(.398)(.142)}$$
$$.345 - .24 \le r_a \le .345 + .24$$
$$.105 \le r_a \le .585.$$

If $C = 0$, then

$$\frac{R_e^2 + R_s^2 - \Sigma(\beta_{e,i}-\beta_{s,i})(r_{e,i}-r_{s,i})}{2} = r_a$$
$$\frac{.602 + .858 - .774}{2} = r_a$$
$$.345 = r_a.$$

To find the value of C, use Formula 30:

$$C = \left(r_a - \frac{R_e^2 + R_s^2}{2} + \frac{\Sigma(\beta_{e,i}-\beta_{s,i})(r_{e,i}-r_{s,i})}{2}\right)\frac{1}{\sqrt{(1-R_e^2)(1-R_s^2)}}$$
$$= \left(.316 - \frac{.602 + .858}{2} + \frac{.774}{2}\right)\frac{1}{\sqrt{(.398)(.142)}}$$
$$= -.12.$$

---

[a] Hypothetical clinician who matched cue utilization coefficients with ecological validity coefficients.

* With $df = 1$, the obtained $\chi^2 = 1.89$, $p < .20$.

equations were obtained for each clinician by relating his estimations of IQ to the Rorschach factors.

Table 5 presents Todd's data (for 39 patients) reviewed in terms of Case VI. Case VI, it will be recalled, is similar to Case V except that the cues are intercorrelated. Also, in this case $R_e^2 \neq 1.00$. The differences between a clinician's beta weights and those for the environment are shown as $(\beta_{e,i} - \beta_{s,i})$; the differences between a clinician's cue utilization coefficients and the environmental validity coefficients are shown as $(r_{e,i} - r_{s,i})$. (Because the beta weight differences were large in these data, the limits of accuracy given by Formula 29 were too large to be of any interest for some of the clinicians. Therefore, Formula 27a was used, which resulted in the limits shown in Table 5.)

"Clinician No. 11" is a hypothetical clinician who matched response process-form with environmental function-form ($R_e^2 = R_s^2$) and utilization coefficients with validity coefficients ($d_\beta d_r = 0$). The data from Clinician No. 11 are used to show that, even if *complete* matching occurs in an environment where $R_e^2 = .60$, achievement can range from .20 to 1.00. Exactly where achievement will fall, within this range, will be entirely dependent upon the value of C, the correlation between the residual variances.

Note that all 10 clinicians listed in Table 5 tended toward greater linearity than actually existed in the environment; the multiple correlation coefficient for each clinician is higher than that for the environment. This implies overweighting of the assumption of a linear, additive environmental function form, an overemphasis which makes the upper limit of possible achievement for each clinician less than 1.00.

But there is also evidence that assumptions aside from linearity affect the judgments of the clinicians, for there is some correlation between the residual variances in their judgments and the residual variance in the (Rorschach-Wechsler) environment. These correlations suggest that the clinicians allow the *pattern* of Rorschach factors to guide their judgments of IQ to some extent. Note, however, that for 7 out of 10 clinicians this correlation is *negative*. Therefore, the relation between the pattern of Rorschach factors and estimates of IQ tends to be the reverse of what it should be. In these cases, the clinicians would have done better had they not considered the pattern effects of Rorschach factors at all.

For example, Clinician No. 6, while attributing too much linearity to the situation (thus restricting his achievement) had a positive C (+.30). This increased his achievement. On the other hand, Clinician No. 8, who also attributed too much linearity to the environment had a large *negative* C (−.43) which decreased his achievement. The ideal is approached by Clinician No. 10, whose R^2 is closest to that in the environment, and who has most nearly matched cue utilization coefficients with environmental validity coefficients. This gave No. 10 the highest achievement limit of any of the clinicians, but his C of only .17 kept his achievement considerably below its upper limit.

In an ecology of this type, then, where linear relationships account for only little more than half of the variance (i.e., $R_e^2 = .60$), achievement depends partly upon perceiving the importance of individual cues, and partly on correctly perceiving the pattern effects of cues. Such pattern effects may be completely unknown to the experimenter, and they may also be inexpressible by the judge. However, if they are present, they may have a positive *or* negative effect on achievement.

The significance of the above analysis is that it permits the investigator to separate achievement into that portion due to pattern effects and that due to component effects.[7]

REFERENCES

AZUMA, H., & CRONBACH, L. J. Performance on a concept-attainment task with scaled cues. University of Illinois, 1961. (Mimeo)

BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272.

BRUNSWIK, E. The conceptual framework of psychology. In R. Carnap & C. Morris (Eds.), *International encyclopedia of unified science.* Chicago: Univer. Chicago Press, 1952.

BRUNSWIK, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

CRONBACH, L. J. Processes affecting scores on "Understanding of others" and "Assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 380–491.

HAMMOND, K. R. Probabilistic functioning and the clinical method. *Psychol. Rev.*, 1955, **62**, 255–262.

HOFFMAN, P. J. The paramorphic representation of clinical judgment. *Psychol. Bull.*, 1960, **57**, 116–131.

HULL, C. L. The problem of intervening variables in molar behavior theory. *Psychol. Rev.*, 1943, **50**, 273–291.

LEE, JOAN C., & TUCKER, R. B. An investigation of clinical judgment: A study in method. *J. abnorm. soc. Psychol.*, 1962, **64**, 272–280.

LEWIN, K. Defining the "field at a given time." *Psychol. Rev.*, 1943, **50**, 292–310.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1962.

MEEHL, P. The cognitive activity of the clinician. *Amer. Psychologist*, 1960, **15**, 19–27.

NEWTON, J. R. Judgment and feedback in a quasi-clinical situation. *J. abnorm. soc. Psychol.* (in press).

POSTMAN, L., & TOLMAN, E. C. Brunswik's probabilistic functionalism. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959.

RAPPOPORT, L. Multiple probability learning and personality. Unpublished master's thesis, University of Colorado, 1961.

SARBIN, T. R., TAFT, R., & BAILEY, D. E. *Clinical inference and cognitive theory.* New York: Holt-Rinehart, 1960.

SIDDIQUI, M. M. Tests for regression coefficients when errors are correlated. *Ann. math. Statist.*, 1960, **31**, 929–938.

SIMON, H. A. Rational choice and the structure of the environment. *Psychol. Rev.*, 1956, **63**, 129–138.

SMEDSLUND, J. *Multiple probability learning.* Oslo: Akademisk Forlag, 1955.

SUMMERS, S. A. The learning of responses to multiple weighted cues. *J. exp. Psychol.*, 1962, **64**, 29–34.

TODD, F. J. A methodological analysis of clinical judgment. Unpublished doctoral dissertation, University of Colorado, 1954.

(Received September 6, 1962)

APPENDIX

PROOF OF SCHWARZ'S INEQUALITY FOR FINITE n

Let

$$(r_{e,i} - r_{s,i}) = X_i$$

$$(\beta_{e,i} - \beta_{s,i}) = Y_i$$

then (A):

$$\left(\sum_{i=1}^{n} X_i Y_i\right)^2 \leq \sum_{i=1}^{n} X_i^2 \sum_{j=1}^{n} Y_j^2$$

is the same as

$$(r_{e,i} - r_{s,i}, \beta_{e,i} - \beta_{s,i})^2 \leq d_r^2 d_\beta^2.$$

To show that (A) is true, use induction on n.

First, for $n = 1$, we have

$$(X_1 Y_1)^2 \leq X_1^2 Y_1^2.$$

---

[7] A FORTRAN program for computing (for each successive trial) the statistics discussed in this article is available at the Behavior Research Laboratory at the University of Colorado.

Assume for $n = k - 1$

$$(\sum_{i=1}^{k-1} X_iY_i)^2 \leq \sum_{i=1}^{k-1} X_i^2 \sum_{j=1}^{k-1} Y_j^2$$

then

$$(\sum_{i=1}^{k} X_iY_i)^2 = (\sum_{i=1}^{k-1} X_iY_i)^2$$

$$+ 2\sum_{i=1}^{k-1} X_iY_iX_kY_k + X_k^2Y_k^2$$

$$\leq \sum_{i=1}^{k-1} X_i^2 \sum_{j=1}^{k-1} Y_j^2$$

$$+ 2\sum_{i=1}^{k-1} X_iY_iX_kY_k + X_k^2Y_k^2 = (\mathrm{B}).$$

Since

$|2ab| \leq a^2 + b^2$, let $a = X_iY_k$ and

$$b = X_kY_i$$

to get

$$2X_iY_kY_iX_k \leq X_k^2Y_i^2 + Y_i^2X_i^2.$$

Therefore

$$2\sum_{i=1}^{k-1} X_iY_kY_iX_k \leq \sum_{i=1}^{k-1} (X_k^2Y_i^2 + Y_k^2X_i^2)$$

so

$$(\mathrm{B}) \leq \sum_{i=1}^{k} X_i^2 \sum_{j=1}^{k} Y_j^2$$

implies

$$(\sum_{i=1}^{k} X_iY_i)^2 \leq \sum_{i=1}^{k} X_i^2 \sum_{j=1}^{k} Y_j^2$$

which implies

$$\sum_{i=1}^{k} X_iY_i \leq +\sqrt{\sum_{i=1}^{k} X_i^2 \sum_{j=1}^{k} Y_j^2}$$

Psychological Review
1964, Vol. 71, No. 1, 61–76

AN INQUIRY INTO THE DOCTRINE OF REMOTE ASSOCIATIONS[1]

NORMAN J. SLAMECKA

University of Vermont

The essential validity of the doctrine of remote associations was assessed. The findings of three original derived-list experiments supported a hypothesis of perception of list patterning, and were incompatible with predictions based upon remote associations. A fourth experiment supported the hypothesis that the association method produces its results through an artifact of serial position. Findings based upon the method of anticipatory and preservative errors were explained consistently by an alternative rationale that excluded remote associations. It was concluded that the doctrine of remote associations is of doubtful validity, and that serial memorization involves acquisition of the items per se, and then the learning of their positions in the list.

In his now classic monograph on rote verbal learning and retention, Ebbinghaus (1885) established the concept of remote associations. His investigations led him to believe that during the memorization of a serial list of items, functional associative bonds were established not only between immediately adjacent items but also directly, and at the same time, between items farther separated from each other in the list. These latter direct connections between noncontiguous items were designated as remote associations. They presumably arose automatically as part of the ordinary process of serial memorization, with no special effort on the part of the subject to bring them about. They were conceptualized as being formed both in the forward and backward directions, that is, an item in the middle of the list could develop a forward connection with the last item and a backward connection with one of the first items. As memorization of the series proceeded, it was thought that each and every item tended to form such a connection with every item in the list, thereby resulting in a complex bundle of associative interconnections among all the individual members of the series, as depicted in Figure 1. Numbers represent the sequential positions of the items as they occur in the list and arrows indicate associative bonds, with remote forward asociations lying above, and remote backward lying below the item sequence.

Furthermore, the strength of any given associative bond was asserted to be a negative function of its degree of remoteness. The bond between Items 1 and 5, above, represents a third degree of remoteness (since it spans three intervening items), and it would be of lesser strength than the bond from 1 to 3, a first degree of remote-

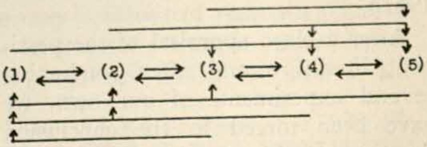


FIG. 1. Schematic illustration of associative bonds among the items of a memorized serial list, according to the concept of Ebbinghaus.

---

[1] This work was aided by National Science Foundation Grant No. G-14, 721. The author is grateful for the assistance rendered by Ernest Lemay and Muriel Whalen.

ness (since it spans only one intervening item). Also, all backward associations were considered to be much weaker than their counterparts in the forward direction. Thus, the remote backward linkage from 4 to 1, above, would be weaker than the remote forward linkage from 1 to 4. To summarize our description of the essential properties of his concept, here is a statement from Ebbinghaus (Woodworth & Schlosberg, 1954):

> With repetition of the syllable series not only are the individual terms associated with their immediate sequents but connections are also established between each term and several of those which follow it. . . . The strength of these connections decreases with the number of the intervening (syllables). . . . Certain connections of the members are . . . actually formed in a reverse as well as in a forward direction. . . . The strength . . . was however considerably less for the reverse connections than for the forward ones [p. 710].

Since its inception some 77 years ago this formulation has remained, for all intents and purposes, essentially unchanged and has been accorded rather solid and widespread recognition in the conceptual armamentarium of the verbal learning theorist. Contemporary reference works in the experimental area discuss remote associations as an important part of the phenomena of serial learning (Hovland, 1951; McGeoch & Irion, 1952; Woodworth & Schlosberg, 1954), and recognition of the concept is also given in several introductory psychology texts (Geldard, 1962; Kimble, 1956; Morgan, 1961).

After a close appraisal of the pertinent literature and after conducting several experiments of our own, we have been forced to the conclusion that such widespread acceptance is definitely ill-advised, since there is much room for reasonable doubt about the validity of the remote associations doctrine (in our estimation it has indeed become a doctrine), and that it serves no necessary purpose in the conceptualization of serial verbal learning.

Experimental data from the remote associations literature have usually been regarded as proof that there indeed are such functional connections between each item and all others in a serial list. However, it will be demonstrated that certain necessary deductions from the remote associations position are not borne out empirically, and that these data are susceptible of other equally (if not more) plausible interpretations. Further, it will be shown that phenomena previously attributed to the necessary influence of remote associations can be produced in the absence of any serial learning whatever. We shall present our supporting evidence in the sections to follow, sections delimited on the basis of the three types of experimental methods which have been utilized in the study of remote associations.

METHOD OF DERIVED LISTS

This method, originated by Ebbinghaus, provided the sole source of the data which he presented as evidence for remote associations. An original list is memorized, then the same items are arranged in a different serial order and presented to the same subject again for learning. The latter is the derived list, and it usually involves a patterned rearrangement of the items with respect to their original arrangement. A first-order derived list is formed by skipping every other item from the original list, a second-order list skips two items at a time, and so on. If the sequence of appearance of items in the original list is designated as 1, 2, 3, 4, 5, 6, 7, 8, 9, then a first-order derived list would have a sequence of 1, 3, 5, 7, 9, 2, 4, 6, 8, and a second-order list would be 1, 4, 7, 2, 5, 8, 3, 6, 9. The first-order list brings together items which were linked by a first degree remote association in original learning, and the second-order arrangement brings together items which were previously related by a second degree association. Keeping in mind that remote asociations of the lowest orders are strongest, it then follows that a first-order list should be learned faster than one of second-order, and the latter still faster than a scrambled or control list whose items are in a random sequence. Ebbinghaus (1885) tested such a hypothesis and, indeed, found it supported by his data in terms of percentage savings in learning time.

We cannot agree, however, that his experiment proved the existence of remote associations, for at least three reasons. First, Ebbinghaus learned by the method of whole presentation, wherein the entire list was always in view. It is possible that, inadvertently, he formed "remote" associations directly, by attending to nonadjacent items in direct succession. Secondly, he read the lists in a regular rhythmical manner, and such an imposed organization itself probably influenced the ease with which certain associations were formed, as shown by Müller and Schumann (Woodworth & Schlosberg, 1954). Third, and perhaps the more important objection, Ebbinghaus was his own subject. Since he prepared his own lists and was fully aware of the principle of their construction, it is quite possible that such knowledge could have produced the results, in spite of whatever other precautions he did take.

We turn, then, from this pioneering work to the subsequent experimental literature on remote associations using the method of derived lists, and find relatively few references.[2] Cason (1926) found no significant facilitation in the learning of first-order lists and concluded that "when a series of items is well learned and practiced until the recitation is very smooth, the forward associations between the adjacent items are the only associations that are formed [p. 316]." But we are left with some misgivings since the materials used were very long prose passages and a poem, and it is possible that the sequential dependencies and contextual constraints found in such material served to mask any effects of remote associations. A study by Hall (1928) must be ruled out of serious consideration because her peculiar method of constructing derived lists precluded any meaningful conclusions about remote associations. (McGeoch and Irion [1952] are also of this general opinion.) By far the best designed and controlled study was carried out by Lepley (1934), who used first-order lists and comparable control lists of 12 nonsense syllables at six different retention intervals. Unfortunately, in no case was any first-order list easier to learn than its original list, and the same applied to a majority of the control lists as well. Such heavy and consistent negative transfer, resulting in a finding of no savings, renders the meaning of the results inconclusive for our purposes. Lepley (1934) himself wrote: "Apparently, the potent factors at work in the present study are factors of interference of some sort and the only evidence of remote excitation we have results from a lesser degree of interference in the test lists at the end of the thirty minute forgetting interval [p. 16].” Such words are hardly calculated to convince the disinterested reader of the reality of remote associations.

---

[2] Since we are concerned only with derived-list studies in which remote associations were sought, we are not including a discussion of two studies in which the focus was upon adjacent backward associations, even though derived lists were used.

Goldstein (1950), with relatively naive subjects, reported substantial savings over the original list in the learning of both a first-order list and a control list, but failed to find any savings advantage of the former over the latter arrangement, contrary to the results of Ebbinghaus. Apparently, degree of acquaintance with the materials used is a factor in the outcomes to be expected.

We will now describe our own experimental efforts to clarify the question of the meaning of derived list data. One common characteristic of all materials used in previous studies is that, with the exception of the control condition, there was always a rigidly patterned arrangement of items in the derived lists. That is, the ordering of the items in a derived list with respect to their original order was always regular or systematic. It is therefore conceivable that the subject, given a derived list, might actually perceive the orderly relation that it bore to the original (such as skipping every other item), and proceed to “learn” it by (for instance) a process of mediation. This might take the form of uttering aloud every other item, and bridging the gaps by silent utterance of the intervening items. If his retention of the original list was perfect he could “learn” the derived list on the first recitation trial and earn an impressive savings score. If such were the case, his behavior certainly would not constitute proof of remote associative bonds, but only of perceptual alertness. Degrees of remoteness beyond the first would be progressively more difficult to perceive, and thus progressively fewer savings would be the inevitable result. Certain statements from previous studies are provocative in this regard. Cason (1926), who found no savings, wrote: “When the third and last part of the procedure had been completed, the subject was asked a number of questions about the experiment. None of the subjects noticed the way the lists had been constructed . . . [p. 309].” On the other hand, in a report where substantial savings were obtained in a study of adjacent backward associations, Garret and Hartman (1926) commented: “When the reversed lists appeared, every *S* immediately guessed the order of the new arrangement, while the order of the reversed-pairs was at once evident to three *S*s . . . [p. 245].”

Clearly, what is needed is some way of minimizing the possibility that the subject will have such an insight, while at the same time permitting savings due to truly remote associations. There is nothing whatever in the general rationale of the derived list method that requires the use of a regularly patterned list, just so long as it allows remote associations to demonstrate their facilitating effects upon learning. Therefore, a modified type of derived list which is of the first-order *on the average,* should produce just as much savings as the regularly arranged first-order list, but without allowing the subject to benefit from perception of patterning. But if perception of patterning is the critical variable, then the modified first-order list should be learned just as slowly as a control list.[3] Experiment I was designed to test these divergent hypotheses.

---

[3] The importance, in a learning situation, of identifying patterns of events is attested to in a paper by Bruner, Wallach, and Galanter (1959) when they write: “learning and problem-solving may be more profitably viewed as identification of temporarily

TABLE 1

MEAN TRIALS FOR OL AND DERIVED LISTS FOR EXPERIMENTS I, II, AND III

Experiment	OL		Patterned list		OL		Modified list		OL		Control list	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
I	1.06	.23	4.94	3.22	1.06	.23	7.61	3.59	1.06	.23	7.94	2.51
II	20.45	8.22	13.30	4.40	19.95	8.01	12.95	7.31	20.60	6.27	13.55	5.17
III	19.90	4.42	10.90	4.71	20.25	7.32	12.05	3.28	20.80	4.81	14.10	5.02

EXPERIMENT I

A total of 36 students served as subjects as part of the general psychology course requirements. Each subject learned one original list and then three derived lists in immediate succession, all to a criterion of one perfect recitation. The order of the derived lists was counterbalanced across subjects. Learning was by serial anticipation on a memory drum with a 3-second rate of presentation and a 6-second intertrial interval. The original list items were 14 letters of the alphabet in alphabetical sequence, from A to N for half the subjects, and from M to Z for the other half. The three derived lists were: (*a*) a control list with items in haphazard order, (*b*) the usual patterned first-order list with every other item skipped, and (*c*) a modified first-order list with 0, 1, and 2 items skipped in an irregular fashion, but with an overall mean of one skipped item. For the A to N list, it looked like this: A, D, E, G, H, K, M, B, C, F, I, J, L, N, corresponding to a skip sequence of 2, 0, 1, 0, 2, 1, (2), 0, 2, 2, 0, 1, 1. The necessary gap from M to B (indicated by the number in parentheses) also occurred in the regularly patterned arrangement, and is a common feature of derived list construction, in order that all the items be used. All lists retained the same first and last items of the original arrangement. The results presented in terms of mean total acquisition trials are found in the top row of Table 1. Original learning (OL) was 1.06 trials because, oddly enough, two subjects needed two trials each to recite the alphabet, in spite of the fact that all subjects were told the initial letter and that there was an alphabetical sequence. The overall F of 10.43 was significant beyond the .01 point ($df = 2/35$), and Tukey's test showed that the patterned list was more rapidly learned than either the modified first-order, or the control list (.01 level gap = 1.86), with no difference between the latter two lists. Furthermore, 10 subjects learned the patterned list on the very first recitation trial, but this did not occur even once for the other two lists! It seems undeniable that such extremely rapid acquisition must have been the result of perception of the pattern. We feel that these data definitely provide strong support for the perception of patterning hypothesis, and at the same time stand against the remote associations doctrine, since the latter would maintain that the modified list should have produced savings equivalent to those of the patterned list.

The fact that savings were not consistently obtained in all past studies

---

or spatially extended patterns . . . [p. 209]." Also, the well-known "discrimination hypothesis" is based upon the organism's perception of regular sequences of events (Tyler, Wortz, & Bitterman, 1953).

may be attributed to the subject's failure to perceive the particular derived list patterning. The familiar alphabet series used for OL in Experiment I was deliberately chosen in order that it would facilitate rapid perception of the derived list pattern. On the other hand, we would predict that an original list of relatively unfamiliar items, such as nonsense syllables, would not produce facilitation in either type of first-order derived list. Although Ebbinghaus used nonsense syllable lists and reported savings in accord with remote association expectations, it can hardly be said that, for him, those materials were unfamiliar, since he invented them and was fully cognizant of their patterning. A truly naive subject would less likely have insight into the systematic rearrangement used for the patterned list, and consequently would show no savings over the control or modified lists. The remote assocations position would predict that since the original list was learned, remote associations must have been formed, and subsequent savings of equal degree should be obtained for both the patterned and modified lists, over the control. Experiment II tested these hypotheses.

EXPERIMENT II

A total of 60 naive subjects drawn from the same source as before was divided into three groups of 20 subjects each. Each subject learned a single original list to a criterion of two successive correct recitations, followed immediately by acquisition of a single derived list to the same criterion. The same presentation rate and intertrial interval were used as before. The lists were made up of 12 nonsense syllables with m' values from 3.03 to 3.12 (Noble, 1961). Two forms of the original list were made up, using the same syllables but in different order, to lessen inadvertent sequence effects. The three derived lists were of the same types as those used before: (*a*) a control list, (*b*) a patterned first-order list, and (*c*) a modified first-order (on the average) list with the following sequence of skipped syllables: 2, 0, 1, 2, 0 (2), 0, 2, 1, 0, 2. Prior to exposure to the derived list, each subject was told that it consisted of the same items as those in the original list, but in a different order. (The subjects of Experiment I were also told this.) Results of Experiment II are shown in the second row of Table 1. It requires no statistical analysis to conclude that there were no significant differences in the rates of learning among the three derived lists. No subject gave any evidence of having detected the principle of arrangement of List *b*. Goldstein (1950) also found that a patterned list was learned no more rapidly than a control list.

It is clear that simply arranging a derived list in patterned low-order form for learning by a naive subject does not automatically produce more rapid acquisition in comparison to the random or control arrangement. This fact is in contradiction to the specific expectation from the doctrine of remote associations, as is the finding of no savings for the modified list, as well. Some knowledge of the patterning must be present, it would seem, before facilitation can reasonably be expected.

What would the outcome of Experiment II have been if each subject did have knowledge of the principle of construction of his second list? We would predict that such knowledge would most benefit the group having the regularly patterned list, would least benefit the control group (knowing that the list is a haphazard rearrangement does not reduce much uncertainty), and might conceivably benefit the modified list group to some extend (knowing that the next item is either 0, 1, or 2 items away in the original list should reduce some uncertainty). Experiment III was performed to answer this question.

EXPERIMENT III

Another 60 naive subjects were used from the same source as before. Essentially, Experiment II above was repeated, with the identical nonsense syllable lists, the three types of derived lists, and the same presentation conditions and acquisition criteria. But these subjects were given full knowledge of the nature of their respective derived lists. This was achieved by: numbering each syllable on the original list in consecutive order from 1 to 12 and retaining these same numbers when the list appeared in derived form, and informing the subject, just before presentation of the second list, of the precise principle underlying its rearrangement. Results are shown in the bottom row of Table 1. Duncan's Multiple Range Test indicated significantly faster learning of the patterned list than of the control list (Rp3 = 3.02, $df = 57$, $p < .05$). The modified list occupied an intermediate position in mean trials but was not significantly different from either of the other two lists (Rp2 = 2.86, $df = 57$, $p > .05$). Thus, with knowledge of the arrangement principle available to them, the subjects did best on the traditional Ebbinghaus type of list. Without such knowledge, as shown in Experiment II, they did no better than control subjects, favorable "remote" associations notwithstanding. Such a distinction would be foreign to the remote associations position, since it views formation of remote bonds as an *automatic* accompaniment of serial learning.

For purposes of control it was necessary to check that the three types of derived lists used in Experiments II and III were of equal difficulty in and of themselves, apart from any transfer effects from the original learning. Thirty naive subjects, from the same source as before, were each given one of the three list types to learn to a criterion of one errorless trial. Mean trials and subjects *SD*'s for the patterned, modified, and control lists, respectively, were: 18.3 (5.14), 20.1 (5.90), and 19.2 (5.81). The F was .23 ($df = 2/27$), certainly not significant. It can be concluded that the derived lists per se did not differ in difficulty.

We conclude from the above experiments that the derived list method in fact fails to provide convincing support for the doctrine of remote association, and instead, serves to give plausibility to an alternative account based upon the subject's perception of list patterning. We will next consider the second method of studying remote associations.

ASSOCIATION METHOD

This method was introduced by Wohlgemuth (McGeoch & Irion, 1952) and later elaborated by others. After having learned a serial list, the subject is presented with each item individually and required to react with the first response that occurs to him. In order to reduce extraneous responses the subject may be instructed to restrict his associations by responding only with a list item. In any event, the production of an item response which was not adjacent to the stimulus item in OL is taken prima facie as representing a remote association established through serial learning. If Item 2 is the stimulus and the subject responds with Item 6, this is construed as a third degree remote forward association. McGeoch and Irion (1952) appear to have no doubts about the essential superiority and validity of this technique when they assert that:

> Results obtained by the association method provide a more direct picture of the specific remote associations formed, are freer than the method of derived lists from the possible masking influences of associative inhibition, and leave no doubt that forward remote associations occur in relatively large numbers [p. 96].

McGeoch (1936) conducted an experiment using this technique, where, in addition to the associations, he also measured the reaction times of each. His results showed that there was a fairly steady decrease in the absolute number of associations (both backward and forward) as their degree of remoteness increased. Such a finding is in accord with expectations, since more remote associations are presumed to be weaker than less remote ones, and consequently less probable of emission. However, the forward reaction latencies themselves clearly contradicted such a position, since they tended to have shorter durations as they became more remote. Thus there is a nagging inconsistency in the findings as they relate to the doctrine of remote associations.

Later, Raskin and Cook (1937) added a logical refinement to the analysis of data generated by this method, and their paper appears to be the last significant word on the development of the technique. They pointed out that it is improper to tabulate only the absolute number of remote associations given, since the opportunities for their occurrence necessarily lessen as the degree of remoteness increases. A correction for opportunity must be made on the absolute totals. In their experiment they found, in agreement with McGeoch, that the absolute totals showed a steady decline with increasing degrees of remoteness, but when corrected for opportunity there was a rise starting at the fourth degree, resulting in an overall U-shaped function. They attributed this rise to summative effects of both backward and forward remote associations impinging upon the same item. They concluded that:

> remote forward associations are formed in the learning of a series of nonsense syllables, but . . . backward associations, also, are formed during the learning of such a series. As in the case of forward associations, the associations in a backward direction are formed between nonadjacent as well as adjacent syllables . . . [p. 393].

We vigorously dispute this conclusion and challenge the validity of the technique on which it was based, on the grounds that the association method involved a serious artifact which renders the results at least ambiguous with respect to the interpretation to be placed upon them. This flaw consists in completely neglecting the decisive implications of the serial position function. It is a well-established empirical fact that the initial items of a list are most easily learned, the last items next most easily, and those just beyond the middle are the last to be mastered. Since serial learning proceeds by complete trials, wherein the entire list is always run through in succession, it follows that the subject will have differential amounts of practice in giving the correct responses to different items. That is, by the time the list is mastered the subject will have had considerable overlearning on the first items, less so on the last items, and only bare mastery of the middle ones. Now, generalizing from the important work of such authors as Goldiamond and Hawkins (1958) and Underwood and Schultz (1960, Ch. 6), a relationship between the prior frequency of a response and its degree of availability may be stated as follows. Prefamiliarization of a response, or the degree to which it has been previously emitted, determines the probability of its emission in a subsequent free-responding situation. We suggest that the association method embodies the essential aspects of such a prefamiliarization procedure, in that the subject gets differential practice on the various OL items (because of their different serial positions), and is later tested for the relative dominance or availability of his responses. Such a state of affairs should logically, in and of itself, produce the pattern of results commonly reported with the use of the association method, and heretofore singly attributed to the operation of true remote associative bonds. We contend that results such as those of McGeoch (1936) and Raskin and Cook (1937) could be obtained with no serial learning entering into the picture whatever, and that therefore they would not bear critically upon the question of whether remote associations are a valid by-product of serial learning. Experiment IV tested this hypothesis.

EXPERIMENT IV

Fifty-four naive psychology students participated as subjects. Each subject was given a controlled amount of differential practice on six nonsense syllables. The syllables, ranging in *m′* value (Noble, 1961) from 3.51 to 3.53, were: CIV, LAT, KEN, DIS, MOR, HAL. Each syllable was typed on a separate card, and after the deck of cards was well shuffled, the subject went through it, looking carefully at each card, spelling the syllable aloud and then pronouncing it. The frequencies of appearance in the deck of the six syllables were 25, 10, 3, 1, 3, 10. This was an approximation to the general shape of a serial position curve, and the subject went through this deck of 52 cards just once. There were three different arrangements of syllable-frequency pairings in order to control for idiosyncratic properties of any given syllable. Upon completion of this differential practice, the subject was tested by the association technique. In order to avoid the emission of unusable, extraneous responses, the subject was asked to respond with the first syllable that came to mind from those he had seen, but not to repeat the stimulus term he was being given. He was presented with each of the six syllables in haphazard order, one time each. The results are displayed in Table 2.

We may regard the items across the top margin of Table 2 as representing the sequence of items in a serial list which had, respectively, 25, 10, 3, 1, 3, and 10 mean total correct anticipations in OL. Along the left-hand margin are the stimuli presented during the association test, and the cells contain the total frequencies of particular responses given as associates. One diagonal row of cells is blank, since instructions forbade responding with the given stimulus syllable. Entries in the top row were not obtained empirically but were derived from the marginal totals of 54, divided according to the proportions required by the obtained column totals and are not included in the grand totals. The (X) represents what would have been the starting symbol for a serial list, that tells the subject to anticipate item number one. Since we lacked the foresight to include such an item during the association test, we decided, for completeness' sake, to synthesize its effect in the controlled manner described in order to provide one more level of forward remoteness for the presentation in Table 3.

The frequencies within any column

TABLE 2

TOTAL FREQUENCIES OF ASSOCIATIONS OBTAINED IN EXPERIMENT IV

Stimulus items	Response items						Total
	1 Practiced 25 times	2 Practiced 10 times	3 Practiced 3 times	4 Practiced 1 time	5 Practiced 3 times	6 Practiced 10 times	
(X)	15	11	7	4	8	9	
1		21	8	2	9	14	54
2	24		7	4	8	11	54
3	14	9		6	14	11	54
4	13	14	10		9	8	54
5	26	8	6	5		9	54
6	17	13	12	6	6		54
Total	94	65	43	23	46	53	324

in Table 2 display some degree of unaccounted-for variability, but an inspection of the overall column totals shows that the attempt to achieve a serial position function through manipulation of the practice frequencies alone was reasonably successful. By the same token, our original statement relating degree of prefamiliarization to the subsequent degree of response availability was confirmed. Table 3 shows the distribution of these same remote associations arranged in degree of remoteness, according to the customary analysis of association method data. For instance, to obtain the total immediate forward responses (zero degree of remoteness), the frequencies in Table 2 running from the upper left corner down to the right through 15, 21, 7, 6, 9 and 9 were summed and appear as an absolute total of 67 in Table 3. To obtain the immediate backward responses, the frequencies running from the lower right corner of Table 2 up to the left through 6, 5, 10, 9, and 24 were summed and appear as a total of 54 in Table 3. The corrections suggested by Raskin and Cook, and the resulting totals, are also presented. The "absolute totals" columns show steady decreases in the totals of remote associations as degree of remoteness increases, in good agreement with McGeoch (1936) and Raskin and Cook (1937). The

TABLE 3

ABSOLUTE TOTALS AND CORRECTED TOTALS OF ASSOCIATIONS IN EXPERIMENT IV, ACCORDING TO DEGREES OF REMOTENESS

Forward associations				Backward associations			
Degree of remoteness	Absolute totals	Correction factor	Corrected totals	Degree of remoteness	Absolute totals	Correction factor	Corrected totals
0	67	6/6	67	0	54	5/5	54
1	45	6/5	54	1	40	5/4	50
2	28	6/4	42	2	33	5/3	55
3	24	6/3	48	3	39	5/2	97
4	22	6/2	66	4	17	5/1	85
5	9	6/1	54				

"corrected totals" columns also tend to approximate the U-shaped function described by Raskin and Cook.

Since it has now been shown that the overall results of the association method can be brought about without any previous serial learning, it follows that data gathered by that method are not at all conclusive with regard to the question of remote asociations, and constitute no clear proof of their existence. Although the responses obtained in Experiment IV may be said to be associations to the stimulus items, they are certainly not the result of any *serial* learning procedure, and are not, therefore, remote associations in the true meaning of the term.

The last of the three methods used in the study of remote associations is discussed below.

METHOD OF ANTICIPATORY AND PERSEVERATIVE ERRORS

When one attempts to recite a serial list from memory, he may omit a few items and give a response further ahead in the sequence, or he may repeat an item that occurred earlier in the sequence. These responses are customarily designated, respectively, as anticipatory and perseverative errors. The method of anticipatory and perseverative errors consists essentially in analyzing the pattern of such errors committed during serial memorization. It has been presumed, rather uncritically, that such errors represent genuine remote associations formed by serial learning. The apparent advantages of this method, according to McGeoch and Irion, (1952), are:

> The anticipatory error method has certain advantages over the two other methods. The remote associations, whatever their degree of remoteness, are overt and specific, which is not the case under the method of derived lists, though it is by the association method. Remote associations may be observed as a function of amount of practice in a single group of subjects. This cannot be done with the method of derived lists, and can be done with the association method only if an association test is made after each practice trial or after selected blocks of trials [p. 98].

Studies by Lumley (1932), Mitchell (1934), and Bugelski (1950) are representative of the use of this method, and their major findings may be summarized as follows: (*a*) anticipatory errors are much more frequent than perseverative errors, (*b*) frequency of such errors is inversely related to their degree of remoteness, even when corrected for opportunity of occurrence, and (*c*) as learning progresses, the number of far remote associations declines both absolutely and in proportion to the number of near remote assocations.

In order to verify such findings from some of our own data, we present in Table 4 an analysis of the OL performance, up to one perfect trial, of the 60 subjects from Experiment II, above. Our data do indeed show the same pattern as those of previous reports: (*a*) out of a total of 1077 remote errors, only 62, or 5.8%, are in the backward direction; (*b*) there are fewer errors the higher their degree, even after correction; and (*c*) as fourths of learning increase, far remote errors drop off faster than near remotes. There is even a tendency, as Lumley (1932) observed, for near remotes to increase somewhat, at least up to the second fourth of learning.

However, as may by now be suspected, our view of the meaning of data such as these is quite different from the usual interpretations that have been put upon them in the past. We choose, in order to test the strength of the traditional doctrine, to reject the assumption that these responses are true remote associations,

TABLE 4

TOTAL NUMBER OF REMOTE ASSOCIATIONS (FORWARD AND BACKWARD COMBINED) ACCORDING TO DEGREE OF REMOTENESS FOR EACH FOURTH OF LEARNING, FROM OL OF EXPERIMENT II

Fourths of learning	Degree of remoteness									Total
	1	2	3	4	5	6	7	8	9	
1st	110	67	40	31	18	11	7	9	3	296
2nd	188	86	26	24	11	2	5	2	1	345
3rd	160	61	15	14	5	1	1	2	2	261
4th	128	27	14	4	1	1	0	0	0	175
Total	586	241	95	73	35	15	13	13	6	1077
Corrected total	586	265	116	100	55	28	29	36	22	

according to the sense of that concept. Simply because the subject makes an intralist error does not necessarily mean that it represents an association that has been built up specifically through *serial* learning any more than if he were asked to utter the first number between 1 and 10 that occurs to him when the stimulus "go" is given. It might be a response made prepotent through much prior practice of it alone. The presumption that such errors indicate serially learned, *direct* connections between spatially and temporally separated items is gratuitous and highly questionable. Rather, they can be more simply and compellingly explained as being the joint result of serial position effects and educated guesses on the part of the subject in the following manner.

As he learns, the subject must acquire the responses themselves (the "response-learning" phase of Underwood and Schultz, 1960), as well as their *relative* positions in the list (somewhat akin to the "hookup" phase of Underwood and Schultz). During the early stage of acquisition the subject has probably mastered only the first few (and possibly the last) items, thus reflecting the serial position influence, and has but a very sketchy grasp on the identity of the remaining items. There is, for him, an uncertain region in the central area of the list, the order of whose items has as yet not been fixed. However, the subject does his best to follow instructions and occasionally makes a guess as to the next item, uttering any one he happens to know except those he has just seen (therefore, very few backward associations), and those (probably the first ones) whose identity and position have been fixed. At this early stage he is more likely to emit remote associations of higher degree than later, since his ignorance of the list is maximal, and his guesses are less constrained. As learning proceeds, the uncertain region of the list shrinks in extent, leaving fewer unplaced items to be guessed, thus logically insuring that the resulting total errors, as well as their degrees of remoteness, will be smaller. It can be seen in Table 4 that the error total for the second fourth of learning is higher than for the first fourth, and yet, according to the account thus far, it should be lower. This may readily be understood on the presumption that the response-learning phase has reached completion during the second fourth of learning, and therefore the subject has more responses available to him for potential guesses than he has during the first fourth.

The preceding rationale predicts the obtained overall pattern of errors in Table 4 and in those of other studies of this type, without drawing upon any dubious assumptions concerning the development of any remote associations. Such a rationale seems to us more parsimonious and somehow more promising than its alternative. For instance, it would be forced to predict that the distribution of the actual serial position sources of intralist errors would show a tendency, as learning proceeded, toward a greater concentration beyond the middle of the list, because of the progressive shrinking of the uncertain region. In other words, the variability of these errors should show a steady decline across successive fourths of learning, as acquisition progressed. The remote associations doctrine would, of necessity, be silent on this question, since it could deduce nothing specific from its premises alone.

A test of this prediction was made from the OL data (up to a criterion of one perfect trial) of Experiment II. The variances of the distributions of intralist errors from the first to the last fourths of learning were, respectively: 7.29, 6.07, 5.20, and 3.91. Thus, the predicted decline in variability was obtained, and the mean location in the list from which the errors came was about the eighth item, which lies precisely in the region last to be mastered. That part of our hypothesis which asserts that the subject guesses at items he has not yet learned, but refrains from naming items that have already gone by, has received empirical confirmation from a recent study by Peterson, Brewer, and Bertucco (1963). Those authors tested the possibility

> that the occurrence of remote associations is strongly influenced by a guessing strategy on the part of the Ss. In the ideal case, after the S had learned what the responses were, if he could remember what items had already appeared on a given trial, then he could confine his guesses to the remaining items [p. 258].

Their experimental results supported such a hypothesis.

At this point we are convinced that data generated by the method of anticipatory and perseverative errors yield more fully to the above stated rationale than to the alternative account invoking remote functional connections.

DISCUSSION

The major conclusion to be drawn is that the classical doctrine of remote associations is doubtful, in the sense that the data upon which the doctrine was dependent do not in fact unequivocally support it, but rather tend to support more strongly other explanatory hypotheses. These other hypotheses must be regarded, at the very least, as being strong alternatives to the classical doctrine in the light of the experiments described above. Nevertheless, what heuristic benefits has the classical doctrine been able to achieve? The main noteworthy use that has been made of it has been its application toward an explanation of the serial position function. Thus, Hull (1935) used it when he spoke in terms of trace-conditioned responses, and later Bugelski (1950), when he made direct use of the method of anticipatory and perseverative errors in developing a serial position rationale. We must maintain, however, that since the premise upon which such explanations rest (that is, the asumption of the existence of true remote associations) has been shown to be open to question it follows that such explanations themselves are not highly likely to survive future investigation. Judging by the evidence already marshalled above, it might be more prudent to say that the serial position function goes a long way toward explaining data usually attributed to the operation of remote associations, and not the other way around. Whatever the ultimately accepted explanation of the serial position curve will be (and papers are still being published about that problem), is is unlikely that remote associations will play any part in it.

What more positive statements can be made at this time about the conceptualization of serial memorization? Certainly, we feel that remote associations are out of the question, but it would be tempting at least to agree that adjacent connections between items are formed. However, even such an assumption cannot be made with confidence, since recent developments have rendered it highly questionable. Young (1961, 1962) has presented a great deal of convincing data to the effect that the stimulus for an item in a serial list is not the preceding item at all, but that it may instead be the serial position of that item. Also, Ebenholtz (1963) has demonstrated clearly that a serial list *can* be learned quite satisfactorily without the necessity of forming adjacent sequential connections in the list, but rather through associating the items with their spatial positions. With these developments to guide us, the most promising (and challenging) statement we can offer to describe serial learning is that it involves acquiring of the items per se, and then the fixing of their relative positions in the list, through associating them, not with each other, but with a self-generated sequential or spatial symbol (such as first, second, etc.). Both of these processes probably overlap in time, with the response-learning stage being initiated first, and the position-learning stage starting shortly thereafter, in much the same manner as the two phases discussed by Underwood and Schultz (1960). It should be emphasized that this position does not deny the existence of associative bonds, but it does deny that serial learning proceeds by a process that results in simple chaining whereby one item is the stimulus for the next, and it also denies the necessity for postulating any remote associations. Rather, the associations formed are more likely to be between the item and some distinctive symbol designating its relative position in the list, and are the result of an active searching process on the part of the subject. An apposite summarization of the serial learning situation was made many years ago by Woodworth (1938):

> We may well be struck also with the importance of perception or apprehension in the learning process—perception of relations, patterns and meanings. To look at a list of number or nonsense syllables, you would think that the thing to be done was to forge links between the adjacent terms, but the actual learning proceeds largely in quite another way. It does not start with elements and unite these, but it starts with groups, or even with the whole series, and proceeds largely by analysis and the finding of parts and relations [p. 35].

SUMMARY

Previous studies relating to the doctrine of remote associations were reviewed, and several new experiments testing the validity of the concept were presented.

The overall results of Experiments I, II, and III served to support an

hypothesis based upon perception of derived-list patterning, and were essentially incompatible with predictions based on the doctrine of remote associations.

Experiment IV gave support to a hypothesis that the association method produces its results because of an artifact, namely, differential practice on the correct responses because of the serial positions of list items. It was concluded that the association method was totally inconclusive with regard to the question of the reality of remote associations.

Findings based upon the use of the method of anticipatory and perseverative errors were consistently accounted for by a rationale based on serial position effects and the subject's guessing pattern, and a further prediction based on that rationale was confirmed.

It was concluded that the doctrine of remote associations is doubtful, and that other alternative hypotheses merit serious consideration. It was suggested that serial memorization involves the acquiring of the items per se, and then the fixing of their positions in the list through associating them with self-generated sequential or spatial symbols.

REFERENCES

BRUNER, J. S., WALLACH, M. A., & GALANTER, E. H. The identification of recurrent regularity. *Amer. J. Psychol.*, 1959, **72**, 200–209.

BUGELSKI, B. R. A remote association explanation of the relative difficulty of learning nonsense syllables in a serial list. *J. exp. Psychol.*, 1950, **40**, 336–348.

CASON, H. Specific serial learning; a study of remote forward association. *J. exp. Psychol.*, 1926, **9**, 299–324.

EBBINGHAUS, H. Memory: A contribution to experimental psychology. (Trans. by H. A. Ruger, & C. E. Bussenius.) New York: Teachers College, Columbia University, 1913.

EBENHOLTZ, S. M. Position mediated transfer between serial learning and a spatial discrimination task. *J. exp. Psychol.*, 1963, **65**, 603–608.

GARRETT, H. E., & HARTMAN, G. W. An experiment on backward association in learning. *Amer. J. Psychol.*, 1926, **37**, 241–246.

GELDARD, F. A. *Fundamentals of psychology.* New York: Wiley, 1962.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuch: The log relationship between work-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

GOLDSTEIN, N. Saving in the learning of scrambled lists as influenced by degree of familiarity with nonsense syllables. *J. gen. Psychol.*, 1950, **42**, 87–96.

HALL, M. E. Remote associative tendencies in serial learning. *J. exp. Psychol.*, 1928, **11**, 65–76.

HOVLAND, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

HULL, C. L. The conflicting psychologies of learning: A way out. *Psychol. Rev.*, 1935, **42**, 491–516.

KIMBLE, G. A. *Principles of general psychology.* New York: Ronald Press, 1956.

LEPLEY, W. M. Serial reactions considered as conditioned reactions. *Psychol. Monogr.*, 1934, **46**,(1, Whole, No. 205).

LUMLEY, E. H. Anticipation as a factor in serial and maze learning. *J. exp. Psychol.*, 1932, **15**, 331–342.

MCGEOCH, J. A. The direction and extent of intra-serial associations at recall. *Amer. J. Psychol.*, 1936, **48**, 221–245.

MCGEOCH, J. A., & IRION, A. L. The psychology of human learning. (2nd ed.) New York: Longmans, Green, 1952.

MITCHELL, M. B. Anticipatory place-skipping tendencies in the memorization of numbers. *Amer. J. Psychol.*, 1934, **46**, 80–91.

MORGAN, C. T. *Introduction to psychology.* New York: McGraw-Hill, 1961.

NOBLE, C. E. Measurements of association value (a), rated associations (a′), and scaled meaningfulness (m′) for the 2100 cvc combinations of the English alphabet. *Psychol. Rep.*, 1961, Monograph Supplement 3-V8, 487–521.

PETERSON, L. R., BREWER, C. L., & BERTUCCO, R. A guessing strategy with the anticipation technique. *J. exp. Psychol.*, 1963, **65**, 258–264.

RASKIN, E. & COOK, S. W. The strength and direction of associations formed in the learning of nonsense syllables. *J. exp. Psychol.*, 1937, **20**, 381–395.

TYLER, D. W., WORTZ, E. C., & BITTERMAN, M. E. The effect of random and alternating partial reinforcement on resistance to extinction in the rat. *Amer. J. Psychol.*, 1953, **66**, 57–65.

UNDERWOOD, B. J., & SCHULTZ, R. W. Meaningfulness and verbal learning. New York: Lippincott, 1960.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

YOUNG, R. K. The stimulus in serial verbal learning. *Amer. J. Psychol.*, 1961, **74**, 517–528.

YOUNG, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *J. exp. Psychol.*, 1962, **63**, 307–313.

(Received December 10, 1962)

Psychological Review
1964, Vol. 71, No. 1, 77–79

THEORETICAL NOTES

CURARE DRUGS AND TOTAL PARALYSIS

KENDON SMITH

University of North Carolina, Greensboro

Question is raised concerning the efficacy of curare drugs in producing total striate paralysis. It is pointed out that early Es reported muscular activity under deep curarization; and that thinking, quite possibly muscular in nature, is unhampered by curarization. It is further suggested that the "exophthalmic response," conditionable under curare, be more fully elucidated as a nonstriate response. Implications are drawn for recent experimentation in the area of conditioning under curare.

Two recent articles by Solomon and his co-workers (Black, Carlson, & Solomon, 1962; Solomon & Turner, 1962) have reviewed the history of curare drugs in the study of learning. The same articles have also provided accounts of the authors' own research in this area. The historical reviews and the experimental contributions are, I believe, thorough and valuable. At the same time, however, I should like to describe a reservation which might be attached to both of them.

The conclusions reached by the papers cited depend completely upon the assumption that curare and its variants produce "total paralysis," that "there can be no skeletal response" in the curarized animal. Indeed, this is a view coming to be generally held (e.g., Munn, 1961, p. 491). It is the essence of the reservation lodged here, nevertheless, that there are good reasons for caution in subscribing to such a view.

The first of these reasons was pointed out several years ago (Smith, 1954): the pioneer experimenters in the area of learning under curare were quite explicit about residual muscular responses even in deeply curarized animals. As Solomon and Turner (1962) themselves remark of the early work of Girden and Culler (1937):

> while the subject [a dog] was totally curarized . . . the US did elicit a twitch of the semitendinosus muscle, even though it did not produce a limb movement. A good semitendinosus muscle twitch CR developed in response to presentation of the CS under curare [p. 205].

Girden and Culler's (1937) own report goes into considerable additional detail about the exact effects of curare upon neuromuscular activity. Thus, they point out that,

> bursts of twitching or "shivering" would ripple over the body; this exemplifies the strychnine-like effect of the drug on the spinal cord, and is not a symptom of inadequate curarization [p. 266].

Weak stimulation of the motor roots of the spinal cord produced in curarized muscle a response which "was definitely *stronger* [p. 271]" than normal. As Girden and Culler finally point out,

> The curare effect seems to consist more in depressing the reaction to strong stimuli than in raising the limen for weak stimuli. Curare exhibits a kind of "hobble-effect"—the muscle can react "thus far and no farther," whatever magnitude of stimulus is employed; but its response to liminal or near-liminal stimuli is but little affected . . . Indeed the strychnine-effect seems to enhance the stimulating value of a near-liminal shock on the motor-root [pp. 271–272].

Girden (1942) later worked extensively with conditioned responses in the *striate* musculature of dogs more or less deeply under the effects of either curare or erythroidine (one of the first of the synthetic curares). In fairness, it must be said that, although Girden reported successful conditioning even in thoroughly drugged animals, he did remark that, "When the animal is in the very deep state, no overt response other than the direct contraction of muscle to shock . . . occurs. At this point, of course, striate CR cannot be evoked at all [pp. 116–117]"; but it must be further pointed out that Girden was working now with the gross behavior of intact animals rather than with the closely observed responses of single muscles.[1]

Strictly speaking, of course, these findings reflect only upon curare itself and upon one of its early variants. A second reason for caution lies, however, in the suggestion by later, independent evidence that even the modern versions of the drug have merely a "hobble-effect" upon the striate musculature. The evidence in point comes from the rather dramatic demonstration carried out by Smith, Brown, Toman, and Goodman (1947). As is now well known, Smith allowed himself, in this demonstration, to be placed deeply under the effects of d-tubocurarine; and appropriate tests revealed "no changes . . . in the electroencephalogram, consciousness and sensorium, or in any aspect of higher central nervous system function [p. 13]." One way of interpreting these findings, and the way to which Solomon and his coworkers (Solomon & Turner, 1962) would undoubtedly be partial, would be this:

> Inasmuch as thought processes are seen to go on unhampered in a totally paralyzed individual, thinking cannot depend upon the function of the striate musculature; and a motor theory of thinking is thus rendered untenable.

Another, and, it seems to me, at least equally defensible way of interpreting the same findings would be this:

> Inasmuch as thought processes are seen to go on unhampered in a deeply curarized individual, and inasmuch as there exists a fair body of evidence favoring a motor theory of thinking, it would appear that even radical curarization does not totally abolish response in striate muscles.

It will be seen that the latter view fits in nicely with the evidence from the earlier animal studies.

A final question is raised by the work of Black, Carlson, and Solomon (1962) themselves. Using both d-tubocurarine and another variant of curare, "Flaxedil," and working with dogs as subjects, they were able to condition what they have called the "exophthalmic response"—an outward bulging of the eyeball. Since this bulging is undoubtedly a muscular effect, and since the only generally known muscles which might cause the effect are the striate extrinsic muscles of the eye, the immediate suggestion, again, is that at least some responsivity remains in curarized striate muscles. Solomon has indicated in personal communication[2] that a species of smooth muscle lying behind the eyeball is responsible for the exophthalmic effect, and such a muscle does indeed seem to exist (Kronfield, 1943, p. 35; White, Smithwick, & Simeone, 1952, p. 252). The substantive issue is thus resolved, but the exophthalmic response would appear to deserve a more complete public explanation than it has received to date.

The qualification which these considerations suggest is plain. It may be much more likely than first appears that the somatic nervous system and the skeletal musculature are importantly involved in conditioning under curare and under other drugs related to it. We are not yet forced to share the view expressed by Solomon and Turner when they remark:

> *we can definitely maintain that differential peripheral skeletal responses (and their associ-*

---

[1] The present writer had occasion, at about this time, to use curare in suppressing the activity of the intra-aural muscles of the cat (Smith, 1943). The drug was helpful; but, even in the deeply curarized animal, totally unresponsive by the usual tests and being maintained by artificial respiration, residual activity in the tensor tympani and stapedius muscles was clearly evident in the variability of the electrical response of the cochlea.

[2] R. L. Solomon, personal communication, March 18, 1962.

ated feedback) could not have been a factor in mediating the discrimination learning which we have clearly demonstrated. Thus, those theories of transfer of training and of thinking which restrict themselves to peripheral skeletal response mediators, and their correlated feedback from the periphery, cannot explain the outcome of our experiment [p. 217; italics in original].

REFERENCES

BLACK, A. H., CARLSON, N. J., & SOLOMON, R. L. Exploratory studies of the conditioning of autonomic responses in curarized dogs. *Psychol. Monogr.*, 1962, **76**(29, Whole No. 548).

GIRDEN, E. Generalized conditioned responses under curare and erythroidine. *J. exp. Psychol.*, 1942, **31**, 105–119.

GIRDEN, E., & CULLER, E. Conditioned responses in curarized striate muscle in dogs. *J. comp. Psychol.*, 1937, **23**, 261–274.

KRONFELD, P. C. *The human eye in anatomical transparencies.* Rochester, N. Y.: Bausch & Lomb Press, 1943.

MUNN, N. L. *Psychology.* (4th ed.), Boston: Houghton Mifflin, 1961.

SMITH, K. Bone conduction during experimental fixation of the stapes. *J. exp. Psychol.*, 1943, **33**, 96–107.

SMITH, K. Conditioning as an artifact. *Psychol. Rev.*, 1954, **61**, 217–225.

SMITH, S. M., BROWN, H. O., TOMAN, J. E. P., & GOODMAN, L. S. The lack of cerebral effects of d-tubocurarine. *Anesthesiology*, 1947, **8**, 1–14.

SOLOMON, R. L., & TURNER, L. H. Discriminative classical conditioning in dogs paralyzed by curare can later control discriminative avoidance responses in the normal state. *Psychol. Rev.*, 1962, **69**, 202–219.

WHITE, J. C., SMITHWICK, R. H., & SIMEONE, F. A. *The autonomic nervous system.* (3rd ed.) New York: Macmillan, 1952.

(Received September 14, 1962)

Psychological Review
1964, Vol. 71, No. 1, 80-85

CARDIAC CONDITIONING AND SKELETAL RESPONDING IN CURARIZED DOGS[1]

A. H. BLACK AND W. M. LANG

McMaster University, Ontario

Smith (1963) has presented "reservations" concerning the appropriate theoretical interpretation of some recent data on conditioning under curare (Black, Carlson, and Solomon, 1962; Solomon and Turner, 1962). These data show, first of all, that Pavlovian conditioning of cardiac, pupillary, and exophthalmic responses occurs in curarized subjects, and secondly, that such conditioning affects subsequent behavior (both autonomic and skeletal) in the normal state. These data were interpreted as indicating that conditioning in curarized subjects "provides evidence to support the position that autonomic conditioning does occur independently of skeletal activity" (Black, Carlson, & Solomon, 1962, p. 25).[2] It was suggested further that,

> Since Pavlovian conditioning procedures under curare seem to affect subsequent skeletal behavior in the normal state, the manner in which this effect is mediated becomes important for all S-R reinforcement theories. . . . We are forced to consider how "information" acquired under curare, when there can be no skeletal response, "carries over" or transfers so as to influence skeletal responses in the normal state. Three possibilities immediately suggest themselves: One is some form of "central" mediation or "pure cognition." Another involves peripheral "mediation" via differential autonomic responding and its correlated, discriminated, feedback stimuli. And finally, combination of these two possibilities might be postulated. (1962, pp. 28–29).

Smith (1954, 1963) argues first, that interpretations such as these are tenable only if *absolutely no* skeletal responding *at all* occurs under curare-like drugs, and secondly, that the curare experiments do not preclude the possibility of skeletal activity which could mediate the conditioned response.

We do not agree with these contentions. Rather, we would argue as follows: (*a*) It is only when the definition of a "skeletal response" is left vague that Smith's criticisms seem valid. When the concept skeletal response is operationally defined, then the curare experiments provide evidence to support the hypothesis that autonomic conditioning, and transfer between curarized and normal states do occur without skeletal responding. (*b*) Even if one assumes that some fragmentary skeletal responding takes place in curarized subjects, one must make assumptions which seem highly unlikely in order to maintain a motor-mediation theory of conditioning and transfer. These arguments are amplified in the following discussion.

1. According to Smith, the hypothesis which we must verify is that "skeletal behavior does not occur in curarized animals." His main point is that if we fail to observe skeletal behavior in curarized subjects, we ought not to "accept" the above hypothesis. Rather, we ought to state simply that we have "failed to reject it" because there may be techniques for determining the presence of skeletal behavior that are more sensitive than the ones which we employed.

One can interpret this criticism in two ways, either as a logical or as an empirical criticism. First, it could be a statement of the impossibility of the absolute verfication of *any* hypothesis. If our experiments fail to show skeletal

---

[1] This work was supported by Research Grant MY-2741 from the National Institute of Mental Health, United States Public Health Service to A. H. Black.

[2] This is not to say that skeletal activity, when it occurs, cannot affect the form of the cardiac response. There is considerable evidence to show that it can. Rather, the point is that skeletal activity is not *necessary* for the occurrence of the cardiac response.

behavior in curarized subjects we cannot conclude that we have *proved* our hypothesis. Interpreted in this manner, the criticism is concerned with the logic of inductive inference, and simply points out that we can never have absolute certainty concerning any hypothesis. In this sense, Smith's criticism is correct.

Secondly, even though experimental results cannot prove our hypothesis absolutely, it seems reasonable to think that they could provide some supportive evidence for it. If this position is accepted, Smith's criticism seems to imply that the evidence already presented is not sufficient to support the hypothesis that skeletal behavior does not occur in curarized subjects. As Smith (1954) has stated in an earlier paper, "Unless an experimenter is meticulous in the elimination of skeletal responses, it will always be possible to account for visceral conditioning by postulating undetected somatic activity [p. 223]." In this second interpretation of Smith's criticism, the basic problem seems to be in determining empirically how to eliminate skeletal responses "meticulously." This can be accomplished only by first operationally defining a skeletal response. Thus, if we could reach agreement as to what operationally constitutes a skeletal response, then presumably, there would be no problem in setting up an experiment which would allow one either to reject or accept the hypothesis that skeletal behavior does not occur in curarized subjects.

In our previous experiments we defined a skeletal response as an overt response; that is, one which was observable by the experimenter in the intact organism. Conditioning of autonomic responses and transfer occurred when skeletal responses, so defined, were prevented. A more stringent operational definition of a skeletal response is provided by the electrical activity recorded from the muscle. We have made observations using this approach as well, recording muscle potentials (EMG) and cardiac responses concomitantly in curarized dogs. A brief description of these experiments follows. Muscle potentials in experimentally naive dogs were recorded from subcutaneous needle electrodes feeding into EEG preamplifiers on a Grass Polygraph. Four locations were used, (*a*) over the biceps femoris on the lateral surface of the left hind limb (L-BF), (*b*) over the biceps femoris on the lateral surface of the right hind limb (R-BF), (*c*) over the sartorius on the dorsolateral surface of the right hind limb (R-S), and, (*d*) on the forehead (F).

The dogs were conditioned in a soundproof chamber. Stimulating and recording equipment were kept in an adjoining room. The CS was a 4,000-cycle tone; the UCS a 4-milliampere electric shock. The CS-UCS interval was 10 seconds, and the shock duration 5 seconds. A delayed conditioning procedure was employed. Curarization was carried out by injecting d-tubocurarine chloride[3] through a polyethelene catheter which had been inserted into the recurrent tarsal vein. A Harvard infusion pump was employed in order to control the rate of injection. D-tubocurarine chloride was injected slowly until respiratory failure occurred. An endotracheal tube was then inserted and artificial respiration begun, using a Harvard Respiration Pump. Drug injections were continued throughout the course of the experiment.

In the typical polygraph records which follow, the top two channels present muscle potential, the third EKG, and the fourth presents a tachographic recording of the heart rate. Also, a signal marker is superimposed on the fourth channel. The tachograph presents each interbeat interval as a vertical distance. The shorter the interval between beats (and the faster the heart rate), the less is the vertical displacement of the pen. The onset of the CS is represented by an arrow above the signal marker line, and the occurrence of UCS is represented by a solid bar on the signal marker line. In order to avoid 60 cycle noise, the muscle potential channels were switched off just before shock presentations.

In some of the following records, slight

---

[3] D-tubocurarine chloride was kindly supplied by E. R. Squibb and Sons of Canada, Limited.

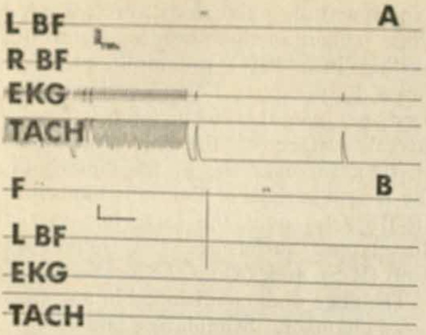


FIG. 1. Heart and muscle potentials following cessation of cardiac activity. (A—Dog N-1. This dog died during the course of the final experimental session. B—Dog S-3. This dog was sacrificed, using Nembutal, following completion of the experiment.)

artifactual perturbations in the EMG tracings were observed. These were of three types. First of all, there were artifacts which were associated with the operation of the equipment controlling stimulus presentations. Secondly, rhythmic oscillations occurred which seemed to be associated with the heart beat. These can be attributed to the movement of the dog and electrodes, produced by the force of the heart beat. This ballistocardiac effect was especially noticeable in curarized dogs. It could be accentuated or decreased by changing the position of the dog with respect to the electrodes and lead-in wires. Finally, there were small artifactual changes in the EMG which seemed to be produced by the recording system itself. These were observed as in Figure 1 where records are shown for two dogs following cessation of the heart beat. Presumably, changes in the EMG at this time would not be attributed to skeletal movements. None of these types of perturbation in the EMG records were scored as skeletal activity.

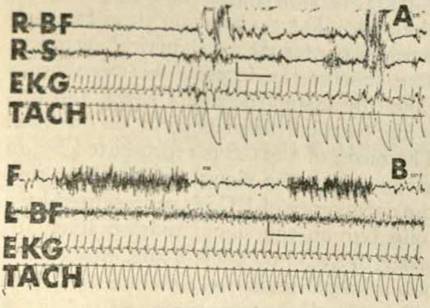


FIG. 2. Heart rate and muscle potentials in the normal state for dog S-4. (See text for explanation of trace identifications. Calibrations: time, 2 seconds; EMG amplitude, 100 microvolts.)

Figures 2 and 3 show data for a typical dog during an experiment on simple Pavlovian conditioning. In Figure 2 the records show EMG and heart rate activity in the normal state before curarization. Figure 3 shows data for the same dog following curarization. Record A of Figure 3 shows the first acquisition trial, and Record B, the sixteenth acquisition trial. On the sixteenth trial there was a clear-cut heart rate response (the difference between the rate during the 10 seconds CS-UCS interval and the 10 seconds preceding CS onset was 33 beats per minute) but no apparent change in EMG. For this dog, cardiac conditioned responses occurred with no detectable muscular movement.

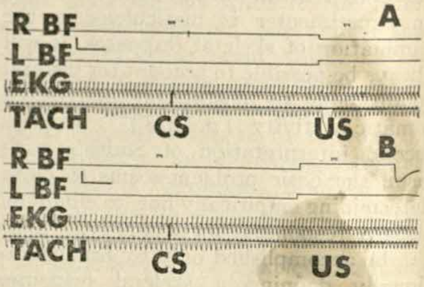


FIG. 3. Heart rate and muscle potentials during Pavlovian conditioning in the curarized state for dog S-4. (A—Acquisition Trial 1. This trial followed the administration of 36 milligrams of d-tubocurarine chloride over 90 minutes. B—Acquisition Trial 16. This trial followed the administration of 102 milligrams of d-tubocurarine chloride over 210 minutes. Calibrations: time, 2 seconds; EMG amplitude, 100 microvolts.)

In a second procedure, one dog was trained in the normal state to avoid shock by flexing its right hind leg. This procedure was designed to increase the probability of skeletal responses to the CS in subsequent conditioning under curare. Data for this dog are presented in Figures 4 and 5. Figure 4 shows the first (Record A), and fifth (Record B) trials on the tenth day of avoidance training in the normal state. When the first trial was conducted (A) there was a considerable number of intertrial responses, but by the fifth trial (B) these had disappeared, and the only large responses that occurred were in the presence of the CS. Following the fifth trial the dog was curarized. Curare was administered slowly in order to observe the gradual decline in skeletal responding to the CS. If any EMG activity was noted in the presence of the CS, the CS was terminated and shock avoided for that trial. If no muscle activity was noted, the CS and UCS were both presented. Record A in Figure 5 shows data for the second trial under curare when some EMG activity was noted. In Record B there seems to be no apparent movement to the CS. There is, however, a very small (approximately 2 beats per minute) but stable[4] cardiac response. Following curarization, the dog was again tested in the normal state and the avoidance response occurred in all 10 test trials.

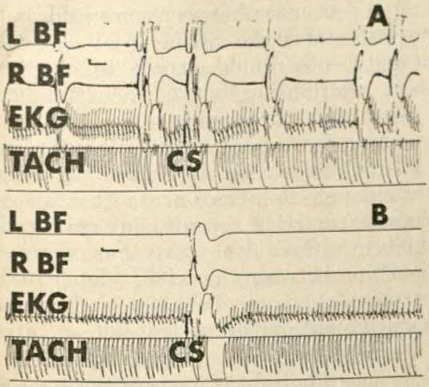


FIG. 4. Heart rate and muscle potentials during avoidance conditioning in the normal state for dog N-1. (A—Acquisition Trial 1 on the tenth day of avoidance conditioning. B—Acquisition Trial 5 on the tenth day of avoidance training. EMG amplitude was reduced in these records because of the considerable activity displayed by the dog. Calibrations: time, 2 seconds; EMG amplitude, 100 microvolts.)

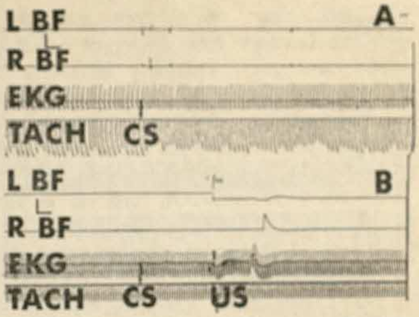


FIG. 5. Heart rate and muscle potentials during conditioning in the curarized state for dog N-1. (A—Trial 2. On this trial some EMG activity occurred and shock was omitted. This trial followed the administration of 18 milligrams of d-tubocurarine chloride over 33 minutes. B—Trial 29. On this trial the CS and UCS were paired. This trial followed the administration of 120 milligrams of d-tubocurarine chloride over 200 minutes. Calibrations: time, 2 seconds; EMG amplitude, 100 microvolts.)

These data seem to support further the hypothesis that cardiac conditioned responses can occur in curarized dogs when skeletal activity, as defined by change in EMG level, is prevented.

These results, of course, do not meet Smith's *logical* criticism. There is always the possibility that some undetected form of skeletal muscle activity could have occurred in this experiment. However, there is one point that should be made here; that is, the same logical arguments apply to the hypotheses involved in a motor-mediation theory of conditioning. For example, even if evidence were provided to show that skeletal responses were found whenever a cardiac conditioned response took place, one could argue that some situation could be discovered in which cardiac conditioned responses would occur without skeletal

---

[4] The difference between pre- and post-CS heart rates averaged between 2 and 4 beats per minute on this series of trials.

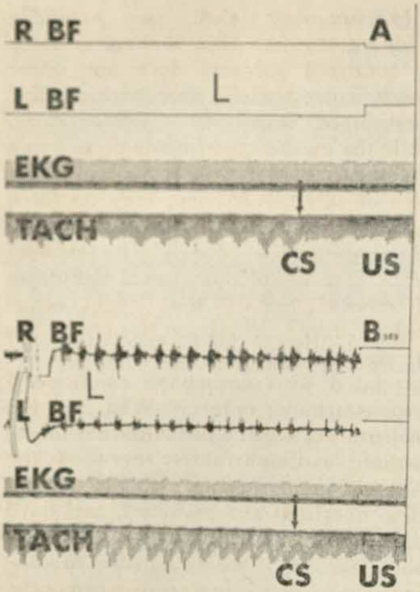


FIG. 6. Heart rate and muscle potentials during Pavlovian Conditioning in dog S-3. (A—A conditioning trial under complete curarization. This trial followed the administration of 42 milligrams of d-tubocurarine chloride over 140 minutes. B—A conditioning trial after the administration of Tensilon which produced partial recovery of skeletal activity. This trial occurred approximately 3 hours following the trial shown in Record A. Calibrations: time, 2 seconds; EMG amplitude, 100 microvolts.)

activity. Thus the logical argument provides no basis for choice between the two positions. Rather, such a choice would have to be made on the basis of the available evidence, incomplete as it may be, and the present evidence does support the hypothesis that autonomic conditioned responses can occur in curarized dogs without being mediated by skeletal activity.

2. Even if one does assume that some undetected form of skeletal behavior occurs in our curarized preparation, the results of our previous experiments, in conjunction with the data described above, still have relevance to motor theories of learning and transfer. Smith (1954, 1963), in his outline of a motor theory of thinking, makes three main points. First, cognitive activity is dependent on skeletal behavior. Under the heading of cognitive activity would be included both the process of conditioning and transfer from the curarized to the normal state. His second point is that skeletal feedback is crucial in the Pavlovian conditioning of autonomic responses; i.e., the only way in which autonomic Pavlovian conditioning can be explained is by assuming that autonomic responses are simply artifacts of the conditioned skeletal responses, or, when the skeletal response is conditioned, then the autonomic response occurs because of innate neural connections relating feedback from the skeletal responses to the autonomic nervous system. His third point is that the cognitive system will be driven adequately even if skeletal responding is reduced, as long as the amount of skeletal responding is greater than zero. This implies that most skeletal behavior is redundant, and that we only need minimal skeletal activity in order to instigate full cognitive activity.

It is at this point that the heuristic deficiency in Smith's system clearly shows itself. If autonomic conditioning and the transfer between curarized and normal states are artifacts of skeletal activity, one would expect to observe some relationship between skeletal behavior and these phenomena. Thus, in order for a motor-mediation theory to be useful we would argue that it is not enough to demonstrate that some skeletal activity occurs in curarized subjects. One also must show a relationship between specific, designated patterns of motor behavior and autonomic conditioning and transfer.

Data which bear on this point are shown in Figure 6. Record A shows a conditioned cardiac response in a typical curarized dog in a simple Pavlovian conditioning experiment. There is an obvious heart rate response to the CS but no apparent EMG response. Three hours after completion of the conditioning phase of the experiment, Tensilon was administered in order to counteract the effects of curare, and further conditioning trials were run. One such trial is shown in Record B. The amount of Tensilon injected permitted partial recovery of skeletal responding. At this stage of recovery, the skeletal activity consisted mainly of bursts of trembling and shivering in phase with the respiratory cycle. This behavior pattern is common among dogs recovering from curare paralysis when body temperature drops.

If there were some relationship between skeletal activity and the cardiac response we would expect a change in the pattern of skeletal activity at the same time, or slightly preceding, the onset of the cardiac response. Furthermore, if skeletal activity were mediating transfer from the curarized to normal state, we would expect *similar patterns* of skeletal activity in the two states when cardiac responses were similar in both states. In the top channel of Record B, skeletal activity is higher during the CS period (especially between bursts of trembling) than during the pre-CS period. This is particularly noticeable during the last few seconds of the CS-UCS interval. Such correlations between cardiac and skeletal responding are common in the fully normal state also. However, there is no such relationship between the skeletal activity and cardiac responses in the curarized state (Record A). In addition, the heart rate response cannot be completely accounted for by correlated changes in muscular activity since there was a cardiac response in Record A very similar to that shown in Record B, but no change in muscular activity in A similar to that in B. Thus, these data provide no clues as to what EMG changes could correlate with cardiac responses under curare, and as to what properties of the skeletal response would act as mediators of the cardiac conditioned response between the curarized and normal states.[5]

In summary, then, our available evidence suggests that skeletal activity in curarized subjects does not occur closely correlated in time with cardiac conditioned responses. Furthermore, while the curare experiments do not rule out the possibility of any conceivable type of skeletal activity, they do force a motor-mediation theory of conditioning to become specific enough to be testable. What is a useful operational definition of skeletal activity that might occur under curare? What are the properties of skeletal behavior which might be correlated with autonomic conditioned responses under curare? What are the properties of skeletal behavior that might mediate cardiac transfer between the curarized and normal states? Until these questions are answered, and until some evidence is provided to support a motor theory, it seems sensible to conclude that autonomic responses can occur without skeletal responding.

REFERENCES

BLACK, A. H., CARLSON, N. J., & SOLOMON, R. L. Exploratory studies of the conditioning of autonomic responses in curarized dogs. *Psychol. Monogr.*, 1962, 76(29, Whole No. 548).

SMITH, K. Conditioning as an artifact. *Psychol. Rev.*, 1954, 61, 217–225.

SMITH, K. Curare drugs and total paralysis. *Psychol. Rev.*, 1963, in press.

SOLOMON, R. L., & TURNER, L. H. Discriminative classical conditioning in dogs paralyzed by curare can later control discriminative avoidance responses in the normal state. *Psychol. Rev.*, 1962, 69, 202–219.

(Received January 30, 1963)

---

[5] While these data suggest that skeletal activity *concomitant* with the cardiac response is not necessary for the occurrence of that cardiac response, they do not preclude the possibility that skeletal responses may be related to cardiac conditioning in some more complex manner. For example, one might argue that skeletal activity during the action of the UCS on at least one or two trials is necessary before conditioning will occur. This hypothesis could not be tested since no EMG records were taken during the action of the UCS. It is not our intention, however, to explore problems such as this in the present paper.

Psychological Review
1964, Vol. 71, No. 1, 86

COMMENT ON THE PAPER BY BLACK AND LANG

KENDON SMITH

University of North Carolina, Greensboro

Brief comment is made on the points and data presented by Black and Lang (1963).

My original note (Smith, 1963), to which Black and Lang (1963) have now replied, concerned itself solely with the question of whether or not curare drugs do in fact forestall responses in skeletal muscles. Even within this limited context, it is possible to comment on one or two of the points raised by Black and Lang.

For example, I have no intention of defending the "logical" criticism of curare experiments—the position that negative evidence, in any quantity, is always inconclusive. We all know that cumulative negative findings are important. It would appear that Black and Lang had begun to develop such findings. If their electromyographic work can demonstrate a consistent blockage of skeletal response by curare drugs, their point clearly will have been made.

Black and Lang are concerned over the very definition of "skeletal response." Their concern has puzzled me, because I have used the term simply to mean "any activity of the skeletal muscles." No doubt the term could be given one "operational definition" or another. It seems much more to the point, however, to accept it at face value and then devise the operations required for detecting skeletal responses so defined. It would appear that the technique of electromyography might be one such operation. The unaided observation of overt response in the intact organism certainly is not.

REFERENCES

Black, A. H., & Lang, W. M. Cardiac conditioning and skeletal responding in curarized dogs. *Psychol. Rev.*, 1964, 71, Pp. 80–85.

Smith, K. Curare drugs and total paralysis. *Psychol. Rev.*, 1964, 71, Pp. 77–79.

(Received February 14, 1963)

the monthly review journal
of the
American Psychological Association

CONTEMPORARY PSYCHOLOGY

reviews

criticism

opinion

on

the latest books, films,

and other instructional media

lists of new books letters-to-the-editor

$1.00 a copy ■ $10.00 a year
($10.50 foreign)

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333-16th Street, N.W.
Washington, D.C. 20036

Bi-monthly

$20.00 a year
($20.50 foreign)
$4.00 a copy

JOURNAL OF COMPARATIVE & PHYSIOLOGICAL PSYCHOLOGY

Editor
W. K. Estes
Stanford University
Associate Editor
Eliot Stellar
University of Pennsylvania
Consulting Editors
Gordon H. Bower
Stanford University
Joseph V. Brady
Walter Reed Army Institute of Research
Byron A. Campbell
Princeton University
Vincent G. Dethier
University of Pennsylvania
Frank A. Geldard
Princeton University
John P. Hall
Pennsylvania State University
Margaret Kuenne Harlow
Wisconsin Regional Primate Research Center
John I. Lacey
Fels Research Institute
Douglas H. Lawrence
Stanford University
Daniel S. Lehrman
Rutgers University
Frank A. Logan
Yale University
Mortimer Mishkin
National Institute of Mental Health
Conrad G. Mueller
Columbia University
Carl Pfaffmann
Brown University
W. N. Schoenfeld
Columbia University
Larry Stein
Wyeth Laboratories
Philip Teitelbaum
University of Pennsylvania
Hans-Lukas Teuber
Massachusetts Institute of Technology
J. M. Warren
Pennsylvania State University
David Zeaman
University of Connecticut

CONTENTS

Weiskrantz, Lawrence, and Cowey, Alan. Striate Cortex Lesions and Visual Acuity of the Rhesus Monkey........ 225

Gross, Charles G. Locomotor Activity Following Lateral Frontal Lesions in Rhesus Monkeys.................... 232

Wilson, William A., Jr., Oscar, Marlene, and Gleitman, Henry. The Effect of Frontal Lesions in Monkeys upon Widely Spaced Delayed-Response Trials................ 237

Warren, J. M., Grant, R., Hara, K., and Leary, R. W. Impaired Learning by Monkeys with Unilateral Lesions in Association Cortex.................................. 241

Stamm, J. S., and Knight, Marcia. Learning of Visual Tasks by Monkeys with Epileptogenic Implants in Temporal Cortex.. 254

Thompson, Robert. Thalamic Structures Critical for Retention of an Avoidance Conditioned Response in Rats...... 261

(continued on inside back cover)

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street, N.W.
Washington, D.C. 20036

VOL. 71, NO. 2 MARCH 1964

PSYCHOLOGICAL REVIEW

FURTHER WORK ON LANGUAGE DEFICIT IN APHASIA[1]

JAMES J. JENKINS AND HILDRED SCHUELL
University of Minnesota *University of Minnesota Medical School*

The writers earlier proposed that language deficit in aphasia could be viewed as a general dimension, and showed that homogeneous scales could be developed with tests selected from an aphasia battery. These scales were re-examined on a population of 157 aphasic patients and again found to meet the usual requirements for homogeneity. The 18 tests used on the scales were examined in the light of a factor analysis of 69 tests in the battery. They were found to load heavily on the 1st principal axis and on the 1st rotated factor, identified as general language. Results were consistent with the hypothesis of a dominant dimension of general language deficit in aphasia.

A few years ago the writers presented an analysis of aphasic test data in support of the hypothesis that there is a dominant dimension of general *language* deficit in aphasia (Schuell & Jenkins, 1959). Twenty-nine tests from the Minnesota Test for Differential Diagnosis of Aphasia (Schuell, 1955), judged to be primarily tests of language performance, were studied to see if a Guttman scale might be developed. Eighteen tests survived an initial screening by phi coefficients, and proved to constitute a homogeneous scale when scored on a pass-fail basis. The scale showed high reproducibility for the total sample of 100 aphasic patients and for diagnostic subgroups treated separately.

Two short scales were constructed, the first consisting of an abbreviated scale of nine tests selected from the 18 for high phi coefficients and wide range of difficulty; the second an a priori clinical scale of seven tests considered to represent an orderly progression in recovery from aphasia. For all patient groups and scale assemblies, the data yielded coefficients of reproducibility that were markedly higher than the minimum reproducibility indices. Cross validation on a small group of patients ($N = 23$) gave further confirmation of the results.

The study was criticized by Jones and Wepman (1961) in connection with their findings on a factor analysis of aphasic test data. Their argument was that restricted selection of tests, the wide range of language impairment represented in the patients, and the method of analysis employed forced the appearance of a Guttman scale where separate language factors could have been found. Schuell and Jenkins (1961) attempted to clarify their position and sought a rapprochement with the findings of Jones and Wepman's analysis, pointing out that differences in the tests employed, pa-

---

[1] This research was supported in part by Public Health Service Grant B-1948, from the National Institutes of Health, United States Public Health Service.

tient samples, and analysis procedures worked to minimize the likelihood that Jones and Wepman would find a general dimension of deficit even if it were present. At the conclusion of the reply the writers attempted to make their position clear with the following statement:

> We realistically expect that, over a large sample of aphasic patients, many dimensions of impairment resulting from brain damage are identifiable, and need to be studied, in addition to the common or general dimension of language deficit. We further expect that, at a given level of language deficit, language tests may be arranged in subgroups which show systematic regularities in aphasic performance in various modalities as well as systematic differences in the performance of various segments of aphasic populations. We do not believe this is in conflict with the notion of a general dimension of language deficit present in aphasia [Schuell & Jenkins, 1961, p. 299].

What we are saying, in short, is that all aphasic patients show a reduction of language that cuts across language modalities. In this respect the aphasic can be compared to an individual who knows a limited number of words, expressions, and permissible sequences in any given language. He will find himself at a loss whether he tries to listen, to converse, to read, or to write. In addition to this general loss, the aphasic may have additional perceptual or sensor-motor disabilities that affect some kinds of language behavior more than others. Luria (1958) has pointed out that it is these special kinds of impairment that make differential diagnosis possible. We are not saying that there is only one clinical aphasic syndrome. We are saying instead that the distinctive feature of aphasia is a reduction of language. Bay [2] makes the same point when he insists upon the necessity for differentiating between aphasia and dysarthria. Our data are here treated to make this point explicit, so that it will not be neglected in the discussion of observed patterns of aphasic impairment and in the discussion of other factors appearing in factor analyses of aphasic performance.

Data have been collected for 157 patients on Form 6 of the Minnesota test, and a factor analysis of 69 tests in the battery has been performed (Schuell, Jenkins, & Carroll, 1962). These new data permit a re-examination of the Guttman scales and at the same time make possible a study of the factorial composition of the tests comprising the scales.

METHOD

Subjects

Subjects were 157 aphasic patients. Of these, 155 were patients hospitalized on the Neurology Service of the Minneapolis Veterans Administration Hospital between June 1955 and June 1958. These subjects were consecutive admissions to the Aphasic Section of the Neurology Service, except for the following exclusions: patients who were neurologically unstable; patients with a psychiatrically confirmed diagnosis of psychosis or regression; patients who responded to no test items; and patients who were evaluated for diagnostic purposes, but were not considered aphasic. In order to increase the number of patients in one diagnostic group, records of two aphasic patients tested by Schuell at Syracuse University, and made available through the courtesy of Louis di Carlo, were included.

Patients were distributed as follows in the diagnostic-prognostic groups: [3]

Group I ($N = 21$), *Simple aphasia without perceptual or motor involvement.* Auditory retention span and auditory recall were impaired with impairment reflected in defective speech, reading, and writing. Speech, reading, and writing improved simultaneously when intensive auditory stimulation was given. These patients made excellent recovery.

---

[2] Bay, E., Symposium on Language Disorders, CIBA Foundation, London, 1963.

[3] These are the same clinical groups as those specified in more detail in the earlier article. The numbering of the groups has been changed as part of the effort to make the classification system more logical and easier to remember. The prognosis for recovery is roughly parallel to the order of the group numbers.

Group II ($N = 27$), *Aphasia with cerebral involvement of visual processes.* Patients much like those in Group I but showing additional impairment of visual discrimination, recognition, and recall. They confused letters and words with similar visual configurations, and tended to spell phonetically. Most Group II patients had visual field defects. Patients recovered speech well but reading and writing improved much more slowly.

Group III ($N = 18$), *Severe aphasia with sensorimotor involvement.* Patients usually showed severe impairment of auditory rentention span and recall. They made consistent articulation errors and speech tended to be labored and hesitant. Speech was functional when acquired but remained limited and defective over long periods.

Group IV ($N = 65$), *Aphasia with some residual language and involvement of both visual and motor processes.* Severity of involvement of each process varied from patient to patient, producing superficially different clinical pictures. There was usually some weakness of the speech musculature. Although a good deal of language was often retained in one modality or another, patients tended not to be capable of sustained or self-directed effort. Functional speech could be increased or made more intelligible in many cases but only limited goals were achieved. Neurological findings were compatible with scattered or generalized brain damage.

Group V ($N = 23$), *Irreversible aphasia with almost complete loss of functional language skills.* Patients in this group had no functional speech, reading, or writing. They made errors pointing to common objects named by the examiner. None of them acquired functional speech, although they usually learned to repeat and copy, and some reactive speech appeared.

Three patients did not appear to be adequately characterized by the above classifications. They showed partial auditory imperception with some residual language intact, although testing revealed rather severe language deficits in all modalities.

Tests

The 18 tests which constituted the previously developed scale are listed below by descriptive titles. More detailed descriptions may be found in Schuell and Jenkins (1959). The tests were given individually as part of the diagnostic testing program on admission to the aphasia clinic.

1. Point to items named
2. Point to letters named
3. Follow spoken directions
4. Listen to sentences, indicate right or wrong
5. Listen to long paragraph, indicate yes or no question on content
6. Match words with pictures
7. Repeat monosyllabic words
8. Complete sentences
9. Name months of year
10. Answer questions (one word)
11. Name objects, pictures
12. Give rhyming words
13. Give biographical information in response to questions
14. Tell three things done during day, and three things a good citizen should do
15. Describe pictured situation
16. State similarity between pairs
17. Write letters dictated in random order
18. Write dictated sentences

For purposes of scaling, each test was treated as a single pass-fail item, and scored as failed if the patient made an error. For the factor analysis the test scores were dichotomized as close to 50% difficulty as possible to permit calculation of tetrachoric correlations.

RESULTS

Plan of Analysis

It was decided to re-examine the largest and the smallest of the scales developed previously, the 18-test scale and the 7-test scale. Each of these scales was tested against the full sample of asphasic patients and against patients in the three diagnostic subgroups separately studied in the earlier paper. It should be noted that this re-examination of the scale does not constitute a bona fide cross validation of the original scaling because 45 of the patients in the present series were mem-

TABLE 1

REPRODUCIBILITY INDICES FOR THE 18-TEST SCALE AND THE CLINICAL SCALE

Scale	Sample	Percent reproducible	Minimum reproducibility
Eighteen test	Total, $N = 157$	88.3	69.7
	Group I, $N = 21$	87.3	71.7
	Group II, $N = 27$	83.1	71.6
	Group IV, $N = 65$	87.5	72.7
Clinical scale	Total, $N = 157$	90.0	69.4
	Group I, $N = 21$	90.5	70.8
	Group II, $N = 27$	86.2	69.9
	Group IV, $N = 65$	88.6	69.6

bers of the original patient sample on which the scales were developed. Thus, 29% of the current group entered into the original determination of reproducibility of the scales. Rather than eliminate this group of patients, it was decided that the maximum sample should be kept so that the comparison with the factor analysis would be as complete as possible.

Next, the tests making up the scales were examined with reference to the findings of the factor analysis of 69 tests from the full aphasic battery for the same sample of patients.

Re-examination of the Scales

Subjects were given total scores consisting of the number of tests they had passed and were ordered from high to low. The tests were arranged in order of difficulty and the distribution of individual passes and fails was plotted. The coefficient of reproducibility for the matrix was then determined.

The results for both scales for the total patient population and the selected diagnostic subgroups are shown in Table 1. Inspection of the table reveals that the reproducibility indices are within a few percentage points of the values found in the first study and, with the exception of the indices for Group II, approximate the usually accepted level taken to demonstrate the existence of a homogeneous scale. As was pointed out in the original study, the lowered efficiency of the scales in Group II may be attributed to the fact that this group has visual losses ranging from mild to severe in addition to language loss, and thus scores more poorly on tests which make visual demands in addition to language demands than the language component of the tests alone would suggest.

Overall, in spite of the complex patterns of visual, perceptual, motor, and sensory motor losses which frequently accompany aphasia, it appears that the severity of language loss itself is scalable, and may be thought of as a relatively homogenous continuum.

Relation of the Scales to Factor Analytic Findings

The factor analysis, reported in detail in Schuell, Jenkins, and Carroll (1962) began with a 69 × 69 matrix of tetrachoric correlations between the tests in the battery. The matrix was subjected to a principal components analysis. Eight principal axes exhausted 99% of the variance to be accounted for. These eight axes were used as the basis for subsequent rotations. Two analytic rotations were performed by the computer, an orthog-

TABLE 2

FACTOR LOADINGS OF 18 TESTS ON PRINCIPAL COMPONENTS AND VARIMAX ANALYSES OF THE APHASIA TEST BATTERY

Test	Percent passing	Principal axes					Varimax factors				
		1	2	3	4	5	1	2	3	4	5
1*	66.9	82	−13	32	−34	10	49	38	40	03	61
2	38.9	91	10	09	07	−06	66	58	19	16	22
3	35.7	91	19	24	02	−14	71	59	16	−01	26
4	33.8	79	−01	−09	−12	−25	59	33	34	21	20
5	14.0	74	13	−05	−04	−14	61	36	19	17	13
6	66.2	81	−25	17	−26	−04	45	35	54	07	43
7*	34.4	87	20	−30	08	02	82	20	22	34	05
8*	44.6	88	38	−08	−14	−03	93	22	17	07	13
9	46.5	90	34	−13	02	09	91	28	19	12	−05
10*	41.4	83	49	−08	−19	00	96	16	06	07	14
11*	34.4	90	34	07	−17	14	91	26	20	00	19
12	19.7	82	24	−31	05	07	84	17	20	22	−07
13	47.1	89	39	00	−12	02	91	29	10	06	18
14*	24.8	87	33	−16	−07	09	90	21	19	16	00
15	19.1	87	38	−09	−16	15	93	17	14	11	16
16*	12.1	88	39	−03	−07	11	91	27	07	08	16
17	23.6	88	−10	13	20	−15	49	73	31	18	10
18	3.2	92	05	12	26	−02	64	68	29	07	−08

* Seven tests on the clinical scale.

onal solution (Varimax) and an oblique solution (Oblimin), and graphical solutions were also obtained. Analyses resulted in virtually identical factor structures showing five major factors and two minor factors identified by only one or two tests. The Oblimin analysis showed the major factors to be moderately intercorrelated with coefficients ranging from .14 to .61.

The factors may be briefly described as follows:

Factor 1. *Language behavior.* Tests involving language functions in all modalities showed high loadings on this factor. The factor accounted for 41% of the variance in the battery.

Factor 2. *Visual discrimination, recognition and recall.* Reading and writing tests obtained high loadings on this factor. In the oblique analysis it was found to correlate most highly with Factor 1 ($r = .61$).

Factor 3. *Visuo-spatial behavior.* This factor involved both visual and sensory processes and correlated most highly with Factor 2 ($r = .55$) and Factor 4 ($r = .58$).

Factor 4. *Gross movement of the speech musculature.* This factor appeared to reflect paralysis or paresis of the speech musculature. It included difficulty in swallowing as well as in imitating movements of tongue and jaw, and palatal involvement.

Factor 5. *Recognition of stimulus equivalence.* This factor was considered to reflect integrative processes involved in taking account of stimulus equivalence, both within and across language modalities.

Overall, it may be seen that a pattern of related factors with a dominant central language factor characterizes the relations between the test scores. This outcome is clearly what would be expected on the hypothesis of a major dimension of language deficit in aphasia.

Table 2 reports the loadings of the 18 tests used in the homogeneous scales on the first five principal axes and on the five major factors on the Varimax rotation.

Study of the table shows that the relation of the scaled tests to the factor structure is somewhat like that found by Gage (1947) in his attitude study. Gage developed a homogeneous scale and performed a factor analysis on a pool of attitude items. He found that items which proved to be scalable were those which had the highest loadings on the first unrotated factor and on the first rotated factor. Items which had lower loadings than the scaled items on his unrotated first factor tended to shift off to other factors in the rotation.

With respect to the principal components in the present study, it is apparent that all of the tests in the scale have very high loadings on the first axis (median loading .87) and very small loadings on the remaining axes. The tests are heavily saturated with whatever is common to the entire battery, as we intended when we selected them for study.

On the Varimax rotation several tests shift away from the general language factor (Factor 1) though all of them have substantial loadings on it. Clearly, the 18-test homogeneous scale is not factorially "pure," but equally clearly, it is very heavily saturated with the rotated first factor which we have called general language. The median loadings of the tests on the five major Varimax factors are .83, .28, .19, .10, and .15, respectively. Fourteen of the 18 tests have their highest individual loadings on the first factor (with loadings from .59 to .96) while four load most heavily elsewhere.

In the entire battery of 69 tests, there are 16 tests which load above .80 on the general language factor. Of these, 10 are included in the 18-test scale and six of the 10 appear in the seven-test clinical scale. Similarly in the entire battery there are 10 tests with loadings of .90 or above on the general language factor. Of these, eight appear in the 18-test scale and five from them appear on the clinical scale.

The four tests on the 18-test scale with the lowest loadings on the general factor appear to be at the ends of the difficulty distribution; they consist of the two easiest tests of the 18, the most difficult test, and the sixth most difficult test. This suggests the reasonable conclusion that the items making up a Guttman scale may well be weak with respect to homogeneity of the scale when they are at the extremes of the distribution, without being detected by the procedures of scale construction. However, the present findings may also be due to the fact that the pass-fail points are different for the scale analysis and the factor analysis, as specified above. Thus, in some of the extreme tests in the scale, we may have been dealing with very poor or excellent performance in which language was most salient while the factor analysis detected other components of performance present in the mid-difficulty range of the test.

DISCUSSION

The results given above indicate that the scales which were developed in the earlier study perform adequately on an expanded population and, furthermore, that the tests in the scales are heavily weighted with the factor of general language as we had hypothesized. It seems clear that these scales may be used as an index of severity of aphasia or the loss of general language ability. The writers much caution, however, against the acceptance of any simple notion of aphasia as varying only in amount. As we pointed out earlier (Schuell & Jenkins, 1959) the degree of general language deficit does not by itself furnish an adequate diagnosis, suggest effective modes of treatment, or predict the extent of recovery. What is shown by these studies is that simple expressive-receptive, motor-sensory, or other elementary classifications of aphasia are inadequate. There is a general language deficit in aphasia which is nonspecific to modalities, tasks or functions. This general deficit must be recognized in any description of the disorder. Students of aphasia since Weisenberg and McBride (1935) have come increasingly to this conclusion, but the level of acceptance of this basic finding by neurologists and psychologists has lagged far behind. Findings of our current research program argue overwhelmingly for recognition of the existence of a dominant general language deficit in aphasia.

Re-examination of the homogeneous scales of language deficit reveals that they perform consistently as expected. When examined in the light of a factor analysis of the entire aphasia test battery, the tests used in the homogeneous scales proved to be heavily saturated with the first principal component of the battery and with the first rotated factor, general language. Tests at the extremes of difficulty of the homogeneous scales appear to be less reliably related to the general language factor obtained on the factor anlysis than tests at the midrange of difficulty. Results are consistent with the hypothesis of a dominant dimension of general language deficit in aphasia.

REFERENCES

GAGE, N. L. Scaling and factorial design in opinion poll analysis. *Purdue Univer. Stud. higher Educ.*, 1947, **61**.

JONES, L. V., & WEPMAN, J. M. Dimensions of language performance in aphasia. *J. speech hear. Res.*, 1961, **4**, 220–232.

LURIA, A. R. Brain disorders and language analysis. *Lang. Speech*, 1958, **1**, 14–34.

SCHUELL, HILDRED. *The Minnesota Test for Differential Diagnosis of Aphasia.* (Rev. ed.) Minneapolis: University of Minnesota Printing Department, 1955.

SCHUELL, HILDRED, & JENKINS, J. J. The nature of language deficit in aphasia. *Psychol. Rev.*, 1959, **66**, 45–67.

SCHUELL, HILDRED, & JENKINS, J. J. Comment on dimensions of language performance in aphasia. *J. speech hear. Res.*, 1961, **4**, 295–299.

SCHUELL, HILDRED, JENKINS, J. J., & CARROLL, J. B. A factor analysis of the Minnesota Test for Differential Diagnosis of Aphasia. *J. speech hear. Res.*, 1962, **5**, 349–369.

WEISENBERG, T., & MCBRIDE, KATHARINE. *Aphasia, a clinical and psychological study.* New York: Commonwealth Fund, 1935.

(Received January 31, 1963)

Psychological Review
1964, Vol. 71, No. 2, 94-103

THE BLANK TRIALS LAW:

THE EQUIVALENCE OF POSITIVE REINFORCEMENT AND NONREINFORCEMENT

M. LEVINE, H. LEITENBERG, AND M. RICHTER

Indiana University

A theorem equating behavior after E says "right" with behavior after blank trials (no outcomes given) was evaluated in 4 experiments: 2-trials-per-problem learning set, modified double alternation, contingent discrimination, and a guessing problem. The theorem accurately described the results in the 4 experiments. The relation of the theorem to a general mediation process model is discussed.

Levine (1963) has recently described a model of hypothesis behavior by humans trying to solve simple discrimination problems. The model compares behavior during two types of problems: Outcome and Nonoutcome problems. An Outcome problem refers to the standard discrimination procedure in which stimuli are presented for a number of trials, the subject must make one of two choice responses after each stimulus presentation, and the experimenter presents an Outcome, e.g., says "right" or "wrong" after each response. A Nonoutcome problem follows exactly the same procedure except that outcomes are not presented, i.e., the experimenter says nothing after each response.[1] For the situation in which a problem lasts for exactly two trials the model yields the theorem that the probability of the subjects repeating their response from Trial 1 to Trial 2 during Nonoutcome problems is equal to the probability of the subjects repeating their response during Outcome problems when the outcome following the first trial is a "right." In mathematical form (letting R = repeated response, N = Nonoutcome problem, O = Outcome problem, $+$ = "Right" after first response):

$$P(R|N) = P(R|O\cap+)$$

Another way of stating this is: *During a Nonoutcome problem the subject behaves as though the experimenter were saying "right."*

Whereas the theorem was derived from a model describing behavior only over the first two trials of a discrimination problem, it is apparent that the theorem might have greater generality. That is, although the model is not developed to describe behavior throughout, e.g., a six-trial Outcome problem, one can still inquire as to the applicability to such a problem of the statement that "during a Nonoutcome problem the subject behaves as though the experimenter were saying 'right.'"

This theorem was evaluated in four experiments. The first, which was intended as a direct test of the model, consisted in a series of discrimination problems each exactly two trials in length. The other three experiments provided a test simply of the generality of the theorem. They consisted in a series of modified double-alternation problems, a series of contingent-discrimination problems, and a single 40-trial guessing problem.

---

[1] It is becoming conventional (Koehler, 1961) to refer to trials without outcomes as "blank" trials. The phrase will be employed in that manner in this paper. A complete problem, however, consisting only of blank trials will be referred to as a Nonoutcome problem.

EXPERIMENT I: TWO-TRIAL DISCRIMINATION PROBLEMS

This experiment has been described in detail elsewhere (Levine, 1963) and will be summarized only briefly here. A learning-set experiment consisting of 90 two-trial problems was presented under four experimental conditions to four groups of 20 subjects. The four conditions differed along a continuum defined by the experimenter's manner of reinforcement. At one end of this continuum is the standard learning-set procedure in which experimenter selects the stimulus to be correct, and reinforces responses only to it on each trial of the problem. This procedure holds for every problem in the experiment. That is, the correct stimulus is the same for both trials of 100% of the problems. The group receiving this condition will be referred to as G-100

At the other end of the continuum is the alternation learning-set procedure in which the stimulus designated as correct by experimenter alternates from Trial 1 to Trial 2.

The correct stimulus is never the same for the two trials of any problem, or, conversely, is the same for 0% of the problems. The group receiving this condition will be referred to as G-0. In between the two extremes experimenter may follow one procedure or the other for as many problems as he wishes. For the two remaining conditions the correct stimulus was the same during 80% of the problems, changing during the remainder (G-80), and the correct stimulus was the same during 20% of the problems, changing during the remainder (G-20). The four groups, then, were G-100, G-80, G-20, and G-0, where the number describes the percentage of problems on which experimenter caused the correct stimulus to perseverate.

Method

Apparatus. The stimuli were 180 different three-letter nonsense syllables selected from Glaze's (1928) lists. Pairs of syllables were randomly selected to form the stimulus objects for a given trial, and were typed .5 inch apart on a 3 × 5 card. A pair of cards containing the same two syllables served as the materials for a problem. A deck of 90 such pairs of cards was constructed.

Procedure. Each subject was shown a sample 3 × 5 card containing two syllables and was instructed that he would receive a deck composed of similar cards, that he was to proceed through the deck one card at a time, and that he was always to choose one of the two syllables on each card. He was further told that the experimenter would say "right" or "wrong" after each choice, and that he was to try to be right as often as possible. These will hereafter be referred to as "standard instructions."

After either 5 or 10 Outcome problems, the experimenter stopped the subject and instructed him that there would now be a test of how much had been learned thus far. The subject was told that during the next few cards the experimenter would not say anything, that because this was a test he was to try to get 100% correct. The subject then proceeded to choose syllables on the next 10 cards (5 problems) with the experimenter saying nothing. After these 5 problems the experimenter announced that the learning would be resumed and presented outcomes during the next 10 problems. Test instructions then followed, etc. Ten Outcome problems continued to alternate with 5 Nonoutcome problems until 90 problems—60 with, 30 without outcomes—had been presented.

Results

The proportion of Nonoutcome problems during which the subjects selected the same syllable on both trials is compared in Table 1 to the

TABLE 1

The Proportion of Problems in which the Responses were the Same on Both Trials

	Group			
	G-0	G-20	G-80	G-100
$P(R\|N)$	.24	.43	.81	.93
$P(R\|O \cap +)$	.28	.40	.83	.92

corresponding proportion during those Outcome problems on which the experimenter said "right" after the first response.

The results in this table are typical of those found by more detailed analyses: there is little difference between the corresponding pairs of proportions. The differences that were obtained were not significant. For the four conditions described by the four groups the theorem is clearly tenable.

Experiment II: Modified Double-Alternation

In this experiment all subjects received a series of four-trial problems, each of which utilized a different pair of stimuli. On any Outcome problem the experimenter nonsystematically selected one of the two stimuli to be correct on Trial 1; on Trials 2 and 3 the other stimulus was always correct; on Trial 4 the stimulus correct on Trial 1 was again correct. In short, the correct stimulus followed the pattern of abba.[2]

On special test problems the group of subjects was subdivided, one subgroup receiving a Nonoutcome problem (experimenter says nothing during all four trials), the other subgroup receiving a (4+) problem (experimenter says "right" after all four trials no matter which of the two responses subject makes). The theorem says that the distribution of obtained response sequences should be identical under these two conditions.

This comparison is the critical test of the theorem. It is possible, however, that the same distribution of response patterns would appear no matter what outcomes were presented. In order to check on this possibility the distribution obtained during a Nonoutcome problem was occasionally compared to the distribution produced when the experimenter said "wrong" on a prescheduled basis.

Method

Subjects. The subjects were 100 students from the introductory psychology courses at Indiana University.

Apparatus. The stimuli were 100 different three-letter nonsense syllables selected from the 80%–93% sections of Glaze's (1928) lists. Pairs of syllables were randomly selected without replacement to form the stimuli for a given problem and were typed on four 3 × 5 cards. The two syllables were typed in capital letters 1.5 inches apart on the horizontal center line of each of the four cards. The positions of the syllables were randomized, with only the restriction that the syllables did not maintain the same positions on all four cards. Four such cards constituted the materials for a problem; a deck of 50 four-card sets constituted the materials for the experiment.

Procedure. The subject received the standard instructions described in Experiment I, was then given the deck, and was instructed to begin. After each trial subject placed the card face down.

---

[2] This is not what is typically meant by "double alternation." This term is used for problems in which the correct stimulus follows an (AABB) pattern. Preliminary work showed that this latter problem series was too easy for college subjects to yield a varied distribution of response sequences. The ABBA pattern is the double-alternation pattern started one step later, hence the adjective "modified."

The correct stimulus on the four trials of each of the first eight problems followed an abba pattern. The syllable defined as A (i.e., correct on Trials 1 and 4 but not on Trials 2 and 3) had been randomly selected in advance. All subjects were treated identically through the eighth problem. On Problem 9 the subject received one of two treatments depending upon the subgroup to which he had been assigned before the experiment began. The 50 subjects of Subgroup 1 now received a (4+) problem. Without any announcement or indication of change in procedure the experimenter said "right" no matter which response the subject made during the four trials of this problem. The 50 subjects of Subgroup 2 received first these instructions:

> For these next few cards I am going to test you on what you have been learning. I want to see if you have learned enough to get the right word without my telling you whether you are right or wrong. This is a test of learning so try to get 100% correct.

They then received a Nonoutcome problem for the ninth problem—the experimenter said nothing during these four trials.

The subjects' procedures were reversed for the tenth problem, i.e., Subgroup 1 received the instructions and the Nonoutcome problem, and Subgroup 2 received the (4+) problem.

The first 10 problems, then, consisted of eight standard Outcome problems of abba solution patterns followed by a (4+) and a Nonoutcome problem on Problems 9 and 10 for one subgroup and the reverse sequence on Problems 9 and 10 for the other. In general, the experiment cosnisted of five such 10-problem series. The first, third, and fifth series ended in a (4+) and Nonoutcome problem; the second and fourth series ended slightly differently. Instead of a (4+) procedure for the ninth (or tenth) problem a (1−, 3+) procedure was employed. In this, experimenter says "wrong" on the first trial and "right" on the subsequent three trials no matter which responses the subject makes.

Results

There are eight response sequences which a subject may manifest on any problem. For example, a subject may single-alternate his selection of syllables for the four trials (i.e., show abab), he may double-alternate (show aabb), perseverate (aaaa), etc. For each of the three types of problems the frequency distribution of the eight response patterns may be plotted and compared. The theorem says that the distribution from the Nonoutcome problems should be the same as that from the (4+) problems. It does not specify the relation of either of these to the distribution from the (1−, 3+) problems, but it implies that the latter should be different. The comparisons are made in Figure 1. The left side of Figure 1 shows the comparison of data from Nonoutcome and (4+) problems from the first, third, and fifth series combined. It will be seen that the match is exceedingly close: the only significant difference is for pattern aaab ($p = .04$ by a t-test for differences between proportions).

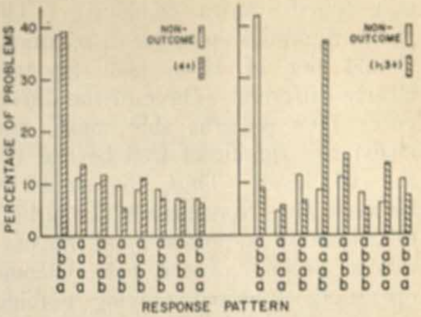


FIG. 1. Distribution of response patterns during (4+) and Nonoutcome problems (left-hand graph), and during (1−, 3+) and Nonoutcome problems (right-hand graph).

The similarity of this pair of distributions is emphasized by the contrast in the other two distributions. The right side of Figure 1 compares the distributions from Nonoutcome and (1−, 3+) problems from the second and fourth series combined. The theorem implies that these distributions should not be alike, and they are clearly different. Three of the differences (for patterns abba, aaab, and abab) are significant well beyond the the .01 level. Thus, the similarity between the Nonoutcome and (4+) distributions is produced by the particular affinity of the two outcome operations involved (saying nothing versus saying "right") rather than by any stereotyped response sets by the subjects.

EXPERIMENT III: CONTINGENT DISCRIMINATION [3]

In the contingent discrimination experiment two stimuli are presented, one of which has been defined as correct by the experimenter. The selection of a stimulus as correct depends (is contingent) upon the value of some other dimension. For example, the background upon which the stimuli are placed may be one of two colors (e.g., red or blue), with the following rule defining correctness: if the background is red then Stimulus A is correct; if the background is blue then Stimulus B is correct. This relationship holds throughout the problem. The subjects received a series of such problems, each problem lasting for exactly four trials.

Method

Subject. The subjects were 120 students from the introductory psychology courses at Indiana University.

Apparatus. Seventy-two different three-letter nonsense syllables (selected from the 80%–93% association sections of Glaze's lists, 1928) formed the stimuli for 36 problems. A problem was constructed by selecting two syllables from this set and typing them on four cards, two of which were one color and two of which were a different color. The cards were selected from six colored sets of 3 × 5 cards (colored blue, green, yellow, orange, red, or white). The two colors employed for a given problem were randomly selected with the restriction that the same color not appear in two successive problems.

The two syllables were typed 1.5 inches apart in capital letters on the horizontal center line of the card. The position of the syllables was randomized with each syllable appearing twice on the left and twice on the right. The card colors either single-alternated or double-alternated over the four-trial problem.

Thirty-six four-trial problems, each utilizing two new nonsense syllables and two colors, were constructed and put into sequences forming a deck of 144 cards.

Procedure. The subject received the standard instructions described in Experiment I. He was then given the deck and was instructed to begin. After each trial the subject placed the card face down.

All subjects received five four-trial Outcome problems in which the correct nonsense syllable was always contingent upon the color of the card. During the sixth problem the subject received one of three different conditions according to the subgroup to which he had been assigned. One condition was a Nonoutcome problem. At the end of Problem 5 these subjects were instructed as follows:

> For these next few cards I am going to test you on what you have been learning. I want to see if you have learned enough to get the right word without my telling you whether you are right or wrong.

Problem 6 was then presented with the experimenter saying nothing. The second condition was a (4+) problem. Without any announcement or change in procedure the experimenter said "right" no matter which response these subjects made during the four trials of Problem 6. The third condition was a (1−, 3+) problem. The experimenter said "wrong" to the first response followed by three "rights" to the following responses no matter which response these subjects made during the four trials of Problem 6.

The first of six problems, then, consisted of five Outcome problems of a contingent discrimination type followed by a test problem which could be either a Nonoutcome, a (4+), or a (1−, 3+) problem. The entire experiment consisted in six cycles of such a six-problem series. Each subject therefore received six test problems. During the first three and again during the last three test problems, he received each of the three test conditions. The conditions were organized for every three subjects to form a Latin square with Tests 1, 2, and 3, and again with Tests 4, 5, and 6.

---

[3] We are grateful to Mymon Goldstein for suggesting this problem as suitable for evaluating the generality of the theorem.

Results

For each of the three types of test problems the frequency distribution of the eight response patterns may be plotted and compared. These comparisons are made in Figure 2. This figure shows three frequency polygons representing the relative frequency distributions of the response patterns for the three conditions. The data are combined for all 120 subjects on the two test problems of a given condition. Each curve, therefore, is based upon 240 problems. It will be seen that as specified by the theorem, the match between the Nonoutcome condition and the (4+) condition is exceedingly close. It is equally clear that, again, the (1−, 3+) condition yields an entirely different type of distribution.

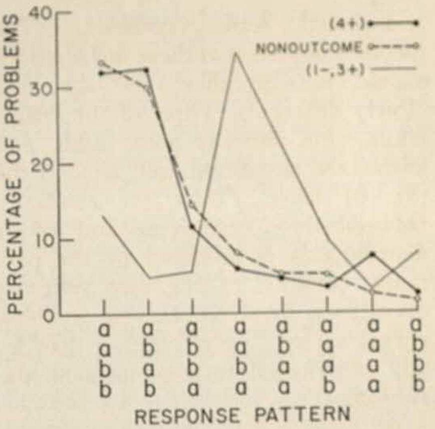


Fig. 2. Percentage of problems on which the eight response sequences occurred during each of the three conditions. (The response sequences are ordered according to the frequency during the Nonoutcome problems.)

Experiment IV: The Guessing Problem

Thus far, the investigations of the theorem have utilized the learning-set procedure, i.e., the presentation of a train of similar problems. Also, sequences of only two (Experiment I) or four (Experiments II and III) test trials were used. It is possible that difficulties might arise with larger numbers of test trials. For example, subjects receiving "blanks" for a longer series of trials may start to manifest less well organized behavior. Or, possibly, subjects receiving all "rights" may, because of the cumulative reinforcing, i.e., strengthening, effect, start to show a stronger tendency toward repetition of the same response. The theorem, of course, specifies that neither "blanks" nor "rights" should have a unique effect or, conversely, that both should yield the same behavior. In order to put the theorem to a more stringent test, as well as to extend its generality, the next application was over 10 trials embedded within a single problem.

The problem selected was a variation of the Humphreys (1939) guessing problem. The subject was required to make one of two responses over a series of 40 trials. The stimulus situation was constant on all trials, i.e., there were no variations in stimuli from trial to trial upon which subjects could make responses contingent. There was, however, a contingency upon the preceding outcome. Two contingencies were employed. In one, the probability that the response correct on

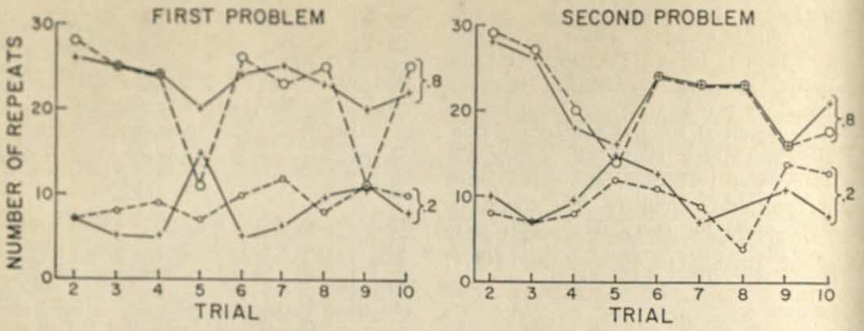


FIG. 3. Number of responses on Trial N which are repeats of the responses made on n-1. (The solid lines are for subjects who were told "right" after each response; the dashed lines are for subjects who were told nothing after each response.)

Trial n was the same as that correct on Trial n-1 was .8; in the other, that probability was .2. The former tended to produce long runs of similar correct events; the latter tended to produce short runs, i.e., alterations, of the correct event.

The subject received one of these procedures for 30 trials. This was followed either by 10 blank trials (experimenter always said nothing) or by (10+) trials (experimenter always said "right" regardless of subject's responses). The patterns of responding obtained under the blank and the (10+) conditions should be comparable, according to the theorem.

Method

Subjects. The 120 subjects of Experiment III also served here. They received the instructions for this experiment following their performance in Experiment III.

Apparatus. Two sets of 40 cards were constructed. In one set the nonsense syllables CED and NIF were typed on all 40 cards in identical fashion (both syllables on the center horizontal line, one inch apart, CED always on the left). In the other set the syllables PUM and DOF were typed in the same manner (PUM always on the left).

Procedures. The subjects were divided into four categories of a 2 × 2 factorial design. The two variables were (*a*) the probability that the same syllable was correct on two successive trials: this probability was either .2 or .8 during the first 30 trials; (*b*) the outcome conditions during the last 10 trials: the subjects received either the blank trials or the (10+) trials during Trials 31–40. Thirty subjects were assigned to each of the four categories.

The subject received the standard instructions described in Experiment I. During Trials 1–30 the experimenter presented outcomes after each response. Right and wrong responses were defined in advance according to the group (.2 or .8) that the subject was in. Within a group the sequence of responses defined as correct was identical for all subjects.

After Trial 30 each subject was told, "The next few trials will be critical trials, the ones I will be chiefly concerned with." The subjects in the (10+) group were then told, "I will continue to tell you whether you are right or wrong"; the subjects in the blank group were told, "I will not be saying 'right' or 'wrong.'"

Following the 10 test trials all subjects were reinstructed for a second problem. The second problem was a replication of the first with only two changes: (*a*) the other deck was used; (*b*) each subject received the other level of both variables. For example, all subjects who had received the .8 and the (10+) conditions on the first problem, received the .2 and the blank conditions on the second problem. All other details of the procedure were identical to the procedure for the first problem.

Results

For Test Trials 2 through 10 the frequency with which the subjects made the same response as on the preceding trial was determined. The results for all four conditions of both problems are presented in Figure 3. It will be seen there that with the exception of a couple of places the blank and (10+) conditions yield very similar results.[4] The overall levels are comparable for each of the four pairs of curves, and the trial-to-trial trends tend to be alike, particularly on the second problem. Also, the magnitude of the differences do not seem any greater toward the end of the 10 trials than at the beginning. That is, there does not appear to be any systematic differential drift produced either by the 10 "rights" or 10 "blanks" alone. In general, the results closely fit the theorem.

One aside is worth recording here. All 120 subjects received the condition of 10 "rights." Not one indicated any suspicions about the experimenter. Most, on subsequent questioning, expressed the idea that they "caught onto the sequence at the end."

DISCUSSION

The finding has been consistently confirmed in four differently structured experiments that the behavior of subjects during trials when no outcomes are given is the same as the behavior of subjects when the experimenter says "right" following each response. Considering the crucial role attributed to "positive reinforcement" for altering response probabilities in contemporary theories of learning it is strange to find that the absence of reinforcement seems to yield exactly the same changes in response probabilities. "Right" has typically been treated as "fixating" a response, as making that response more likely to occur. Why has this effect not appeared in these experiments?

Levine's (1963) model, from which the blank trials theorem was derived, suggests one solution to this problem. One relevant assumption in the model is that the unit of behavior directly affected by the outcome procedure is the subject's hypothesis (a mediating process defined as a prediction about what constitutes solution behavior) rather than his specific choice response. The hypothesis held produces either responses to selected cues (e.g., to nonsense syllables rather than positions in the above experiments) or the sequential organization of choice responses (e.g., alternation, win-stay-lose-shift, etc.) or both. The model also stipulates that each hypothesis is manifested in a specific response sequence, and that both "rights" and blank trials produce no changes in the subject's hypothesis and, therefore, no difference in the sequence of responses manifested. In contrast, "wrong" causes the subject to alter or reject his hypothesis, and thereby produces a very different response series.

It is worth noting that this view, presented here to account for behavior during blank trials, entails two ideas of general significance. One is that the behavioral unit is a mediating process of the sort just described. The other is that "right" functions not as a strengthener of the choice response but as a maintainer of the mediating process. Both are necessary assumptions for treating the results of the experiments described above. Without the first it is difficult to explain the development of alternation learning sets (Experiments I and II), the transfer of these sets to Nonoutcome problems (see Levine, 1963, for a detailed presentation of these arguments), and the curious fluctuations during the 10 test trials of Experiment IV. The second assumption, in conjunction with the first, accounts for the absence drift toward 100% performance of one of the two choice responses when, during the test trials, the experimenter says only "right."

---

[4] Trials 5 and 9 of the .8 condition in the first problem provide the only pairs yielding significant differences ($.01 < p < .05$ for both pairs of points). It should be remembered, however, that 36 comparisons were made. The probability of finding two or more significant differences out of 36 tests is well above .05.

Several human learning models incorporate assumptions sufficiently like these two that with minor modifications they could qualitatively account for the present results. These all invoke a mediating process either directly as such (Kendler & Kendler, 1962; Levine, 1963; Restle, 1962) in terms of selective sampling of cues (Bower & Trabasso, 1963), or in terms of an observing response (Atkinson, 1961). Restle, and Bower and Trabasso, make most exactly the assumption that positive reinforcement maintains the mediating process sampled (as opposed to strengthening one of the two choice responses) but the assumption is closely approximated by the others. The chief modifications necessary in most of these models to account for the blank trials law in its present context are (*a*) the expansion of the mediation concept, which is typically a mechanism for cue selection to handle sequence learning, and (*b*) the addition of the assumption that during blank trials also the subject maintains the same mediating process.

The general theory just sketched accounts for some puzzling findings from other experiments. Buss and Buss (1956), using the Wisconsin Card Sorting Test, compared the effects of saying "right" versus "wrong" for correct and incorrect responses, respectively, with the effects of saying nothing versus "wrong" for these two types of responses. The theorem predicts no differences between the two groups and that is what Buss and Buss found (see also Buss, Braden, Orgel, & Buss, 1956). More striking, however, is the fact that in the latter group almost all the subjects (14 out of 15) reached the criterion of 10 consecutive correct responses with the experimenter never saying a word! According to the theoretical ideas elaborated above, the subjects selected the correct hypothesis at the outset and the correct behavior pattern emerged undeterred. One other finding in this area comes from Buchwald (1959). He has found that once a discrimination has been learned following the usual outcome procedure ("rights" versus "wrongs") the correct response does not extinguish (over 72 trials) when throughout extinction the experimenter says nothing. The blank trials law nicely predicts these results. A mediation process model nicely elucidates them.

REFERENCES

ATKINSON, R. C. The observing response in discrimination learning. *J. exp. Psychol.*, 1961, **62**, 253-262.

BOWER, G., & TRABASSO, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford, Calif.: Stanford Univer. Press, in press.

BUCHWALD, A. M. Extinction after acquisition under different verbal reinforcement combinations. *J. exp. Psychol.*, 1959, **57**, 43-48.

BUSS, A. H., & BUSS, E. H. The effect of verbal reinforcement combinations on conceptual learning. *J. exp. Psychol.*, 1956, **52**, 283-287.

BUSS, A. H., BRADEN, W., ORFEL, A., & BUSS, E. H. Acquisition and extinction with different verbal reinforcement combinations. *J. exp. Psychol.*, 1956, **52**, 288–295.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

HUMPHREYS, L. C. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

KOEHLER, J. Role of instructions in two choice verbal conditioning with contingent partial reinforcement. *J. exp. Psychol.*, 1961, **62**, 122–125.

LEVINE, M. Mediating processes in humans at the outset of discrimination learning. *Psychol. Rev.*, 1963, in press.

RESTLE, F. The selection of strategies in one learning. *Psychol. Rev.*, 1962, **69**, 329–343.

(Received January 29, 1963)

Psychological Review
1964, Vol. 71, No. 2, 104–122

AGGRESSIVE CUES IN AGGRESSIVE BEHAVIOR AND HOSTILITY CATHARSIS [1]

LEONARD BERKOWITZ

University of Wisconsin

Anger arousal creates a readiness to act aggressively but suitable cues—stimuli associated with the present or previous anger instigators—presumably are necessary if aggressive responses are actually to occur. In cases of hostility displacement, as others have proposed, the target evokes aggressive responses from an angered person. Aggressive cues in the angered person's thoughts or in the external situation "activate" an aggressive response sequence which does not attain completion until the anger instigator is injured. Preventing the completion of the activated sequence heightens the arousal state and this (a) increases the strength of any subsequent aggressive responses, (b) raises the level of tension experienced by the person. Completing the ongoing aggressive sequence by inflicting injury often leads, in the absence of guilt or anxiety, to a pleasurable reduction in the experienced tension.

Few observations in the behavioral sciences are as trite as the one boldly proclaiming that an individual's actions are a product of both environmental conditions and internal states. But although we are often told that B = f (P,E), this proposition is sometimes unduly neglected by students of motivation. Even when it is not forgotten, the basic formulation is usually only paraphrased, and we are not given any specific statement as to the manner in which environmental stimuli and internal states interact to determine the direction and intensity of the person's strivings. Such a neglect seems to be particularly characteristic of writers in the field of aggressive behavior. The present paper will attempt to demonstrate that aggression cannot be completely understood without a consideration of both internal conditions and environmental stimuli.

Some preliminary remarks are in order. "Aggression" is here defined as behavior whose goal is the injury of some person or object (cf. Dollard, Doob, Miller, Mowrer, & Sears, 1939), while "anger" is here regarded as the emotional state resulting from a frustration presumably creating a readiness for aggressive acts. In order to simplify this discussion, I might also note, comparatively little attention will be given to the effects of internal controls and inhibitions upon these aggressive responses.

Summarizing briefly, the paper will argue that anger and learned habits separately or together create a readiness to act in a hostile manner, and particular cues—stimuli associated with the present or previous anger instigators—are necessary if aggressive responses are actually to occur. A variety of experimental findings consistent with this thesis will then be described, after which we will consider some of the implications for such behavioral phenomena as displacement and hostility catharsis.

---

[1] The present paper is an expanded version of a talk delivered at the AAAS–Academy of Psychoanalysis symposium, "The role of violence in human behavior," held in Philadelphia in December 1962. The author's research reported here was conducted under grants M1540, National Institute of Mental Health, and G23988, National Science Foundation.

In the usual commentary on hostile behavior the instigation to aggression is pictured as an ever-active energy source constantly impelling aggressive responses. Whether these responses appear in overt behavior, however, supposedly depends primarily on the presence of inhibitions or other avenues of expression, and not on the stimulus qualities of the available targets. Thus, once his aggressive "drive" has been aroused, whether by appropriate biological processes or some frustration, a person presumably is constantly inclined to attack someone—anyone—unless he can displace his urge onto some inanimate object, or divert the drive into some socially acceptable channel.

Hartmann, Kris, and Loewenstein (1949) provide an excellent illustration of this type of thinking. Their revision of Freud's basic doctrine posits an innate biological drive based on unspecified internal sources. Although they do state (almost in passing) that disliked or frustrating objects "invite the discharge" of the aggressive energy, these writers place particular stress on the "assumption of a constant driving power comparable to that of the libido [p. 29]." Like libido, aggressive energy supposedly must find some outlet. The urge may be displaced, sublimated, or neutralized, but after expression it builds up again, and again is said to seek release. The drive object is theoretically relatively unimportant in this process. The object, they maintain (citing Freud), becomes attached to the drive "only in consequence of being peculiarly fitted to provide satisfaction [p. 20]." A particular target is attacked not because it has the stimulus qualities to evoke aggression but because of the satisfactions earned in the course of aggressing against this object.

In one way or another, many discussions of aggressive behavior employ a somewhat similar analysis, regardless of whether they see aggression as instinctual or acquired, the expression of internal energy or a reaction to frustration. As one example, the eminent anthropologist, Clyde Kluckhohn (1945) suggested that there was at least some "freefloating aggression" in every society because of the deprivations and frustrations suffered by the people in it through the course of their lives. This pent-up force constantly seeking expression is said to be drained off in some societies through periodic wars, but other cultures presumably are able to channel the energy into more constructive paths such as literature and public works.[2]

The habitually hostile individual does not have a reservoir of aggressive energy pressing for discharge. Nor, I believe, does he have a year-long accumulation of anger pent up within him. His frequent aggressive outbursts are not emotional explosions from a storehouse of repressed

---

[2] The present analysis fully concurs with Allport's (1954, p. 358) criticism of the "drainage theory" of aggression. Writers influenced by the orthodox Freudian conception of aggression have envisioned a free-floating aggressive energy which may be channeled into many different activities (e.g., attempting to master others) or against a wide variety of objects. Carrying out any of these substitute actions, or attacking any of the substitute objects, then supposedly drains the reservoir of aggressive energy. Contrary to such reasoning, however, there is evidence that hostility toward one group of people is associated positively with enmity toward other groups. The ethnocentric individual, we know today, has learned to dislike a remarkably wide range of people (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). He does not become friendlier toward one outgroup merely as the result of expressing hostility towards other such outgroups. The aggressive action does not in itself weaken his habitual modes of conduct toward strangers or people who are different from him.

I do not mean to question the psychoanalytic notion of hostility displacement. Much of my recent research has been concerned with this phenomenon (e.g., Berkowitz, 1962; Berkowitz & Green, 1962), and I have little doubt that it exists. My argument is with another aspect of the above conceptions. Typically, these views of aggression assign a very minor role to stimulation from the external environment. Such stimulation theoretically affects only whether the person believes aggression is safe or morally justified. For example, in the usual discussion of the scapegoat theory of prejudice (e.g., Williams, 1947) the victim is supposedly attacked primarily because he is a visible and safe target for pent-up hostility within the prejudiced individual. He is a "conveniently chosen outlet," in the words of Hartmann and his colleagues, who is selected largely because the attacker anticipates little or no punishment in return from himself or others. To put this formulation in simpler terms, the aggressive drive supposedly "pushes" aggressive acts unto whatever target happens to be available and is safe to attack. My own position is somewhat different: Basically, I suggest that a target with appropriate stimulus qualities "pulls" (evokes) aggressive responses from a person who is ready to engage in such actions either because he is angry or because particular stimuli have acquired cue value for aggressive responses from him.[3]

Such a conception has several important implications. For one thing, we do not have to think of a person as being constantly driven to attack someone unless his aggressive impulses can be diverted into socially acceptable actions. His likelihood of engaging in hostile behavior can be substantially lowered, theoretically at least, by removing external stimuli capable of evoking aggressive responses.

Similar Conceptions in Other Fields

The position just outlined is similar to the ethological concept of *releaser.* A releaser, or sign stimulus, is a cue in the external environment which produces a given reaction from an organism ready to make such a response. To illustrate, Tinbergen (1951) has reported that an aroused male stickleback fish reacts with fighting behavior to the red on the belly of a rival male. This red stimulus evoking the attack is highly specific; aggression does not occur, for example, if the red is on the rival's back. The evoking stimulus, furthermore, has little if any effect

---

energy or anger but, rather, are reactions evoked by appropriate cues. There may well be no aggressive responses, either overt or internal and hidden, if he did not encounter such stimuli. What this means, then, is that his customary aggressive behavior actually is not revealed unless there are relevant stimuli in the situation confronting him. Moreover, it is not necessary to provide substitute activities in order to lessen the likelihood of aggressive behavior even if anger is occasionally generated by frustrations (the blocking of ongoing goal-directed activity). Unless the thwarted person is kept aroused or is re-aroused by being reminded of his frustrations, his anger probably will dissipate with time, and the probability of aggressive behavior will decline.

[3] In their respective discussions of aggression, both Buss (1961) and the present writer (1962) independently came to emphasize the role of external stimulation in evoking aggressive responses. For Buss anger "energizes" aggressive behavior and lowers the "threshold for the occurrence of aggressive responses [p. 60]." However, contrary to the present formulation, Buss does not seem to regard anger as only a predisposing state, requiring the presence of some cue functioning as a releaser if the aggressive behavior is actually to occur.

if the organism is not in a suitable condition. In the case of stickleback aggression the necessary prerequisite apparently is the production of hormones associated with reproduction. Given the required internal state, the appropriate stimulus releases behavior the organism is ready to make. There clearly is an interaction between the organism's internal condition and the external stimulation. Research has shown that a strong releaser is necessary to elicit a reaction if the animal is under a low arousal condition, while a weaker releaser is sufficient if the animal is in a stronger motivational state (Hess, 1962, p. 185).[4]

Clifford Morgan's (1959) physiological theory of drive also has important similarities to the present conception. According to Morgan, arousal conditions, such as food deprivation or the accumulation of sex hormones in the blood, create a motivational state within the central nervous system (Central Motive State) which predisposes the organisms to react in certain ways to particular stimuli. Although Morgan suggests that some of these motive states may lead to specific patterns of behavior independently of external stimulation, as evidently happens in the case of a female cat in heat, he places particular stress on the "priming" function of the CMS. "Certain reactions are preset or primed by drives," he states, "so that they are released when the appropriate stimulus is present [p. 660]."

For many different kinds of behavior, then, the response depends upon the joint operation of both internal states and external stimulation. Evidence obtained several years ago by Campbell and Sheffield (1953) indicates that even random activity following prolonged food deprivation requires external stimulation. Contrary to the usual simple notion of an internal imbalance driving the organism into activity, rats starved for three days displayed heightened activity only when an environmental change was created. The internal condition alone did not lead to the random behavior.

For more narrowly defined, focused responses rather than general activity, more specific stimuli evidently are necessary. Research into the effects of electrical stimulation of chicken brains (von Holst & von Saint Paul, 1962) suggests that specific situational cues are required in evoking aggressive behavior. In these experiments stimulation of a certain region of the fowl brain led to organized patterns of aggressive behavior primarily when relevant cues ("an enemy, real or artificial") were present. An electrically stimulated rooster would exhibit "only motor restlessness" when "all substitutes for an enemy" were lacking.

In general, a variety of research findings suggest that external stimulation often has a vital role in evoking motivated behavior. Particular internal states may be necessary, but such conditions are not sufficient; in many kinds of behavior overt activity does not occur unless appropriate situational cues are also present to elicit the actions. Some internal emotional states, probably including anger, result only in a general excitement or diffuse motor restlessness in the absence of such evoking stimuli. To go even further, for some kinds of physiological arousal at least, external stimulation can increase the internal "drive" state as well as provide the releasing cues (cf. Malmo, 1962, pp. 402–405).

---

[4] The ethological concepts of "vacuum activity" and "displacement," which attribute certain activities solely to surplus energy within neural centers, appear to contradict the present analysis. However, as the writer has pointed out elsewhere (1962, p. 20), these conceptions have been questioned by various authorities (also cf. Hinde, 1960, pp. 207–208).

Aggression-Evoking Stimuli

Elsewhere (Berkowitz, 1962) I have suggested that an object is capable of evoking aggressive responses to the extent that it is associated with previous anger or aggression instigators. The association is with the immediately preceding frustrater in the case of a person who had just been angered, and probably is with the people from whom aggressive actions had been learned in the case of those with aggressiveness habits.

Such a formula, of course, is only an extension of Neal Miller's (1948) well-known analysis of displacement in terms of stimulus-response generalization. Applying the conditioning principles developed by Pavlov and others, Miller had proposed that tendencies—to make or to avoid making a given response—originally elicited by a particular object generalize to similar objects. The strength of this generalization is in direct ratio to the association between the latter objects and the original instigator. Thus, suppose a man is angered by Person A but cannot attack him because A is absent from the scene. If the man should then encounter other people with varying degrees of association with Person A, these people will elicit hostile reactions from him in proportion to their association with A. That person bearing the strongest association with A will "pull" hostile responses which, although not as strong as those the original instigator would have elicited, are still stronger than the responses evoked by people having weaker associations with A. (We cannot here go into the complications introduced by the presence of inhibitions against aggression. For a discussion of these complexities see Miller, 1948, and Berkowitz, 1962, pp. 107–129.)

The implications of this model should be made explicit. As I said earlier, it maintains that the angered individual is only aroused, i.e, primed, to commit an act of aggression. Whether he does perform such an act depends upon, among other things, the presence of stimuli associated with the anger instigator. Anger intensity does affect aggression. The intensity of the aggressive response evoked by an object probably varies directly with *both* the intensity of the anger and the strength of the association between the object and the original angering agency.

An experiment by Weatherley (1962) provides an excellent demonstration of the interaction of cue and emotional state in aggressive behavior. Female college students whose mothers had been either high or low in permissiveness toward aggression were deliberately angered (in one condition), or were not provoked, and then were given projective test (TAT) cards containing strong or weak cues for aggression. (That is, the scenes depicted on these cards were either strongly or weakly associated with common aggressive situations.) Maternal permissiveness toward aggression probably leads more to relatively weak internal restraints against aggression than to the development of a strong, constantly active aggressive "drive." Thus, it is not surprising that the permissively reared students in the nonaroused condition gave no more aggressive responses to the TAT than did the less permissively trained group in this condition; the two nonaroused groups did not differ in their instigation to make aggressive responses. It was not until the women were provoked that significant findings emerged, but here too, relevant cues were necessary. The angered students whose mothers had permitted aggression exhibited reliably more fantasy aggression than the similarly reared but nonprovoked group only to the strong cue cards. Their aroused hostile inclinations were not revealed, even though their inhibitions were fairly weak, unless aggressive cues were present.

The scenes portrayed on the projective test cards presumably served as aggression-evoking cues to the extent that they depicted situations in which aggression ordinarily occurs. The subject apparently associated the high cue scene with prior situations in which she had behaved aggressively. Because of this tie the card could evoke hostile responses from her fairly easily if she was now angry and ready to make such responses. In the next study the aggression-eliciting stimuli were associated with the immediately preceding arousal condition rather than with common aggressive situations.

In this preliminary experiment[5] (based on only seven subjects in each condition), male college students, taken one at a time, were first given a task by one of the two experimenters. This experimenter was introduced, depending on the condition, either as (*a*) a graduate student in physical education and a college boxer, or (*b*) a graduate student in Speech. Then, for half of the cases given a particular introduction, the experimenter (Experimenter 1) deliberately insulted the subject during the first task, while in the remaining cases he treated the subject in a neutral, nonangering manner. Following this, the second (or neutral) experimenter showed the subject a 7-minute-long film clip, either a fairly brutal prize fight scene or one dealing with the travels of Marco Polo. (We might also note that the subjects seeing the prize fight were given a brief story outline which, according to our earlier findings [Berkowitz & Rawlings, 1963], serves to lower internal restraints against aggression.) At the conclusion of the movie, after the subject had completed a questionnaire assessing his reactions to the film, he was asked to rate each of the experimenters on special forms. Since these questionnaires supposedly were to go to the Dean's Office as part of a survey of student reactions to experiments, unfavorable ratings of the graduate student-experimenters could conceivably hurt them and, therefore, are here taken as acts of aggression. We will be concerned with the two critical questionnaire items asking whether the subject would like to have the given experimenter as a Discussion Section instructor, and whether he would like to be in another experiment conducted by this person. The scores on the two items were summed for each experimenter to give us a single index of expressed unfriendliness toward the experimenter.

Now, what do we expect? The insulting experimenter (Experimenter 1), as source of the subject's anger, should of course receive greater expressions of hostility than the neutral experimenter. But remember also that in some cases Experimenter 1 is fairly strongly associated with an aggressive film; he is seen as a "boxer" by some of the men witnessing a movie about boxing. If this aggressive movie tended to arouse previously learned aggressiveness habits, as I believe is often the case (cf. Berkowitz, 1962, Ch. 9), the aroused hostile inclinations should generalize most strongly to external stimuli bearing the strongest psychological ties to the film. In other

---

[5] The experiment was conducted by the present writer, Frederick DeBoer, and Mark Heironimus.

TABLE 1

MEAN CONDITION SCORES ON THE HOSTILITY INDICES, QUESTIONNAIRE RATINGS OF EACH OF THE TWO EXPERIMENTERS

	Subjects angered by Experimenter 1				Nonangered subjects	
Film:	Boxing movie		Travel movie		Boxing movie	
Identification of Experimenter 1	Boxing	speech	Boxing	speech	Boxing	speech
Experimenter 1	21.42$_a$	15.29$_{bcd}$	16.00$_{bc}$	17.43$_b$	11.00$_e$	10.71$_e$
Experimenter 2	12.86$_{cde}$	9.14$_e$	10.10$_e$	11.71$_{de}$	10.43$_e$	12.14$_{cde}$

Note.—The ratings of the two experimenters were subjected to one "repeated measures" analysis of variance. Cells having a subscript in common are not significantly different at the .05 level. The higher the score the greater the expressed unfriendliness toward the given experimenter. Each mean is based on seven cases.

words, the insulting experimenter should receive the most hostile ratings when he is also relatively strongly associated with the other arousal source: the aggressive prize fight scene, i.e., when he has been labeled a boxer rather than a speech student.

The results summarized in Table 1 indicate that this expectation was fulfilled. The greatest hostility was directed toward Experimenter 1 when he had insulted the subjects and when he had some association with the prize fight movie. Since he was associated with both arousal sources, the insult and the aggressive film, he had a relatively strong cue value for aggressive responses. He had the stimulus qualities which "pulled" fairly intense hostility from the subjects who were primed to make aggressive responses because of the insult they had received.

Much of the experimental research employing notions of the sort described here involves associations based on physical similarities. In the last study, however, Experimenter 1's association with the aggressive film was varied by means of verbal labels applied to him; he was a "boxer" or he was a "Speech student." One of the major difficulties in using associationistic hypotheses to deal with socially important psycholgical problems stems from the great variety of ways in which associations arise. We cannot always specify beforehand just what will be the basis for a psychological tie between various objects or, for that matter, whether a person will regard the objects as being different in important ways. Our thoughts may equate different objects for a number of reasons. We may regard objectively different people as being similar, for example, because we have learned to apply the same label to them (e.g., "They're all Democrats"), or perhaps because they arouse the same emotional feeling in us.

The hostility displacement involved in the scapegoat theory of prejudice can be readily understood in terms of associations created through the aggressor's thoughts (Berkowitz, 1962, Ch. 6; Berkowitz & Green, 1962). Our research at Wisconsin indicates that the hostile responses evoked by one frustrater generalize fairly readily to other disliked people. Although the evidence for a generalization process is not as good as we would want in these studies, we believe that the thwarted person somehow associates his victims with the immediately frustrating agency. All of these people are disliked, for example, and all may be regarded as essentially alike in being unpleasant. An industrial worker thwarted by his employer may displace his resentment onto Jews, then, because he previously had learned to dislike Jews and, disliking them, subjectively associates Jews with the disliked and frustrating employer. He may also associate Jews with the employer because he thinks of all of these people as "businessmen," and association bonds can summate (cf. Berkowitz, 1962, p. 160). However the association arises, an angered person presumably associates the scapegoat group with his more immediate frustrater. Because of this tie, the scapegoat pulls hostile responses from the person who is ready to act aggressively.

Hostility Catharsis

The idea of a hostility catharsis is surely one of the most widely accepted doctrines in the folklore of both the man in the street and the social scientist. For centuries, going back to Aristotle, people have believed that the exhibition of an emotion can "purge" an individual of that emotion. Contemporary formulations of this notion generally deal with anger and aggressive behavior, and generally maintain that a person's hostile actions can reduce the strength of his aggressive tendencies, or somehow make him feel better.

The conception of aggressive behavior outlined so far has some fairly definite implications for the analysis of hostility catharsis. Before going into these implications, however, some basic distinctions must be made, in essence considerably restricting the phenomena included within the scope of the catharsis hypothesis. In the psychological literature, at least, "catharsis" seems to refer most frequently to a lessening in the strength of the *instigation to aggression* supposedly produced by a hostile act. Such a "drive reduction" is not necessarily involved in the second meaning of the term, *a feeling of pleasure or tension reduction* presumably arising from the performance of an aggressive response. Empirical evidence indicates the thwarted person may obtain some tension release by attacking his frustrater (cf. Berkowitz, 1962, Ch. 8; Berkowitz, Green, & Macaulay, 1962), especially if he earlier had wanted to and had been prepared to aggress against that individual but, for some reason, had not been able to do so right away. This tension reduction can be pleasurable, but it is problematic whether the overt hostility will necessarily lessen the likelihood of any further aggression against this frustrater. Even if his aggressive act does lessen the thwarted person's immediate anger, an aggressiveness habit (or negative attitude) may develop; the frustrater may acquire the stimulus properties which, under the appropriate conditions, can cause him to evoke aggressive responses from his victim on some later occasion. In a word, the frustrater may function something like Martin Luther King to a rabid Southern segregationist.

Research into Hostility Catharsis

The present view of aggression as evoked behavior suggests that some additional details should be incorporated into the writer's previously published analysis of hostility catharsis. Summarizing, I have maintained (Berkowitz, 1962, Ch. 8; Berkowitz, Green, & Macaulay, 1962) that much of what is commonly meant by the term "catharsis" can be understood as a special case of a "completion tendency."

The idea of a completion tendency certainly is not new in psychology. Lewin had discussed such a proposition in the 1920's, and two of his students have published supporting data.

Thus, we have Zeigarnik's demonstration that there is better memory for incompleted tasks than for completed ones, and Ovsiankina's study indicating that people are more inclined to resume interrupted activities than to take up tasks again that they had recently completed (Lewin, 1935, pp. 242–244). Failure to complete the tasks presumably created the tension leading to their better recall and the tendency to resume them.

An activity will tend to continue, of course, until the activity goal is reached. Preventing the organism from reaching this goal, i.e., interrupting the activity, can be a source of disturbance. Sheffield (1954) has hypothesized that there is an "increase in excited emotion" when a consummatory response—completing the activity sequence—is stimulated but not allowed to occur completely.

Applying this formulation to aggressive behavior, we would say that inflicting injury on the intended target is the goal or consummatory response completing the aggressive response sequence. *As long as a person is instigated to aggress* he does not obtain "completion" until he sees that he has injured his target, or that someone else has done so.

Tension Reduction Following "Completion." Perhaps the best demonstrations of tension reduction following the completion of an aggressive sequence have been reported by Hokanson. In the first of his experiments (Hokanson & Shetler, 1961), male and female university students were led to believe either that the experimenter was a high status person—in this case, a middle-aged "university professor"—or that he was a low status person—a fellow student. Half of the people in each of these two conditions, while working singly, were deliberately angered by the experimenter, while the remaining subjects received a kindlier treatment from him. The systolic blood pressure obtained at this time showed that physiological tension[6] increased in all of the insulted groups independently of the instigator's status. Following the arousal, half of the subjects in each of the now four conditions were provided with a series of sanctioned opportunities to give electric shocks to the experimenter. Physical aggression against the low status instigator appeared to be tension reducing. The systolic blood pressure in the angered, aggressing students decreased to the level existing in the nonaroused group, but the pressure remained high in those people not shocking their insulting peer. The latter presumably believed they had not injured the instigator and, therefore, did not complete the aggressive sequence.

The results are somewhat puzzling for the high status experimenter condition. In this case, systolic pressure decreased in both the nonaggressing and aggressing angered students to the level within the nonprovoked groups. We might conjecture that the insulted nonaggressing students had quickly given up the idea of attacking the high status professor. They no longer wanted or expected to aggress against him so that, in essence, an instigated aggressive sequence was no longer in operation. If such an interpretation is correct, these findings lend further support to my suggestion that anger arousal in itself does not lead to aggressive responses even if these responses are inhibited.

---

[6] Many neurophysiologists feel that one particular conscious state cannot always be attributed to a given physiological pattern. Thus, high systolic blood pressure is not necessarily indicative of the same conscious emotion in all people. The most parsimonious interpretation is to regard high systolic pressure as indicative of physiological tension.

A second experiment (Hokanson & Burgess, 1962) showed that the findings obtained in the low status experimenter condition are reliable and, in addition highlight the importance of perceived injury to the instigator. Only a low status experimenter was used this time but, as before, half of the subjects, male and female college students, were deliberately angered and half were not. Also as in the earlier experiment, one group of subjects could shock the experimenter in a permissible manner after the provocation, while the others were given a more neutral task. Two other treatments were also carried out. One group created a story in response to a TAT card, and a final condition gave the experimenter questionnaire ratings (i.e., this last condition had an opportunity to administer verbal aggression) after the arousal.

Systolic blood pressure and heart rate measures again indicated that there was a significant physiological arousal following the provocation. The frustrated subjects given an opportunity to injure their tormenter, either physically (by shocks) or verbally (through the ratings), then showed a lessening in physiological tension so that their final scores were not different from those in the nonfrustrated groups. The physiological tension remained high, however, in the insulted TAT and neutral response conditions. Assuming he wants to and is prepared to attack his frustrater, the angered individual must believe he has injured that person in retaliation for the injury he has received if there is to be a physiological relaxation. Fantasy aggression in response to the TAT card in this study did not provide "closure." Fantasy aggression in general, I believe, is relatively unlikely to be satisfying, i.e., produce completion.

Before proceeding further, let us again note the important proviso: *there is no completion* (in the sense used in this paper) *unless there has been an ongoing aggressive sequence.* What are the conditions determining whether such an "activated" response chain exists? Consider the case of Person A who has been angered by Person B. *The anger arousal in itself,* I contended, *does not necessarily give rise to an ongoing aggressive sequence.* Appropriate cues, stimuli associated with the anger instigator, presumably must also be present if the aggressive responses are to be made. Once such responses occur (perhaps only implicitly in his thoughts), Person A theoretically will not obtain completion until he believes that B has been aggressively injured.

However, if the aggressive sequence is set into operation, but completion is prevented, internal tension is induced which is channeled into whatever response happens to be under way at the time. (This last proposition is consistent with views offered by many psychologists—e.g., Amsel, 1958; Brown & Farber, 1951; Mowrer, 1960; Sears, Whiting, Nowlis, & Sears, 1953.) Extending the proposition to aggressive behavior, we have the frustration-aggression notion that blocking aggressive reactions increases the predisposition to further aggression.

An experiment by Thibaut and Coules (1952) provides data supporting this latter prediction. In the first part of their paper they reported that male college students who were provoked by an insulting message from a peer and then were given an opportunity to communicate back to him had less residual hostility toward the instigator at the end of the session than did a similarly angered group not provided with this communication opportunity. However, the investigators were well aware that they had not really demonstrated the existence of a hostility catharsis. The condition differences could be due to a frustration-induced rise in hostility in the *No Communication* condition. In order to test this possibility they performed a second experiment in which every aroused subject was permitted to reply to the instigator, but the men in one group were interrupted for three minutes before they could avail themselves of this opportunity. Analysis of the messages sent to the frustrater confirmed the frustration-aggression hypothesis. The people who had had to wait before replying sent a reliably greater volume of aggression to the instigator than did the students responding more rapidly to the provacation even though the instigator was not responsible for the interruption.

Heightened Predisposition to Aggression Following Inhibited Aggression. This last finding, of course, is another example of hostility displacement. The aggressive inclination engendered by the experimenter's interruption apparently had been displaced onto the other subject. (Or to put this in other terms, the disliked stimulus person had evoked more intense hostility because of the stronger arousal state within the interrupted person.) However, while such a finding is consistent with the concept of displacement, we usually think of displacement in terms of anxiety-inhibited aggression. That is, the angered person is unwilling to attack his frustrater because of fear or anxiety and then takes out his resentment on a safe, available target.

But are these two examples really different if we consider only the blocked aggressive sequence? The anxiety-induced inhibitions are also frustrating and, like the experimenter's interruption in the Thibaut-Coules study, should also lead to a heightened arousal state. If an angered person cannot attack his frustrater right away as he wants and is set to do because of his inhibitions, he is thwarted still more, and, as a consequence, there is a stronger predisposition to aggress against other available targets. Where the less inhibited individual directs his aggression primarily against the person whom he believes had provoked him and exhibits relatively little displaced hostility toward other objects, the restrained individual shows comparatively little overt aggression toward the frustrating agent he cannot or dares not attack—but relatively strong open hostility toward other objects. The aggression against these latter substitute targets has been heightened by the thwarting of the direct attack upon the anger source.

Results obtained in one of our Wisconsin studies of movie violence demonstrate such a displacement process at work (Berkowitz, Corwin, & Heironimus, 1963). Male college students, again taken one at a time, were deliberately angered by a graduate student experimenter in one condition, while the remaining students were given a neutral treatment. Following this brief interaction, two-thirds of the subjects in each of these conditions watched the filmed prize fight scene I had mentioned earlier, and the other subjects saw an equally long movie about English canal boats. The prize fight subjects were given one of two kinds of introductions to the film. Half of them were given a story outline more or less justifying aggression, while the synopsis provided the other subjects in this "aggression movie" condition did not sanction aggression. We have reason to belief that the former movie introduction produced weaker inhibitions against aggression than the latter introduction (cf. Berkowitz & Rawlings, 1963). Finally, all of the subjects completed a questionnaire in which they rated the experimenter and the experiment. For our present purposes we can confine ourselves to the questionnaire results for the angered men witnessing the prize fight movie.

What we can expect if inhibitions against aggression did vary with the movie introduction? The angered subjects having the relatively weak restraints should exhibit comparatively strong hostility toward the source of their anger (the experimenter) on the final questionnaire and a lower level of resentment toward other things, such as the experiment, associated with him. The anger source, we might say, receives more intense aggression than other associated objects. The more inhibited-angered group, on the other hand, would be somewhat less likely to attack the anger source, the insulting experimenter, and comparatively more likely to attack the associated object, the experiment. This was indeed the case. In the less inhibited-angered group 13 of the 15 subjects displayed stronger hostility toward the experimenter than toward the experiment. However, in the angered group presumably having the stronger inhibitions, only 6 of the 15 showed stronger hostility toward the source than toward the associated experiment ($\chi^2 = 7.02$, $p = .01$). Thus the latter were somewhat more likely to displace their resentment onto the experiment rather than the insulting experimenter.[7]

Expectations to Aggress. Use of the frustration-aggression formula requires a clear specification of the nature of frustration. Preventing a person from satisfying a need *we* think he possesses does not necessarily mean the person is actually frustrated. He is frustrated, I have argued in other pages (Berkowitz, 1962, Ch. 2), *when an ongoing response sequence is blocked.* A writer thoroughly engrossed in his work is not frustrated just because he has been without food for a number of hours; there are no ongoing eating response sequences in either the writer's thoughts or his overt activity that are prevented from reaching completion. He is just not concerned with getting food. Similarly, an angered person is not frustrated merely because he has no opportunity to attack the thwarting agency; being angry does not necessarily mean there is an activated or ongoing aggressive response sequence.

This important qualification must be kept in mind in evaluating the results of the Rosenbaum and deCharms (1960) modification of the Thibaut-Coules study. Contrary to the earlier investigators, Rosenbaum and deCharms found that subjects high in self-esteem who were not permitted to aggress against their frustrater were *not* more hostile at the conclusion of the session than were the other high self-esteem men given an opportunity for aggression.[8]

---

[7] In comparison to the mean scores existing in the insulted, less inhibited group, the angered inhibited subjects displayed less open hostility toward the frustrating experimenter but more hostility toward the experiment.

[8] The low self-esteem subjects appeared to show a "catharsis" effect. However, as I have suggested elsewhere (Berkowitz, 1962), the men who were low in self-esteem in this study may well have experienced strong aggression anxiety and, consequently, inhibited signs of open aggression when they were given an opportunity to retaliate because they were not sure how they would be accepted by the other people in the experiment. In other words, they were tempted to aggress, felt anxious because the

Why the difference between these two experiments? Obviously it is not enough just to say an instigated aggressive sequence was prevented from reaching completion in the Thibaut-Coules men while there was no such ongoing sequence in the Rosenbaum-deCharms subjects. We also have to indicate why this difference might have arisen. Examination of the procedures carried out in the two studies suggests a possible explanation for the negative results in the latter experiment which is consistent with the approach taken throughout the present paper. *Unlike the aroused men in the Thibaut-Coules study, the provoked people in the later Rosenbaum-deCharms experiment had not expected to be able to retaliate against their tormenter.* Not anticipating an opportunity to aggress, there conceivably were no ongoing aggressive responses in the latter men. More concretely, since they knew they would not be able to reply to the person who had insulted them, they probably were not thinking of how they were going to tell their tormenter off and, therefore, they conceivably were not disappointed when they could not do so.

A recent experiment from our Wisconsin laboratory was designed to test this analysis.[9] The subjects, male college students, were assembled in pairs, ostensibly for an investigation of problem solving under stress. Each subject worked alone on an assigned problem believing his partner (who was sitting in another room) would give him electric shocks if the partner thought the problem solution was inadequate. In four of the five conditions after the experimenter had collected the written problem solutions and had left the room supposedly to deliver them to the partner, he administered eight shocks to the subject. Since the subject believed they came from his peer in the adjoining room, this excessive number of shocks generally caused the subject to become angry with his partner.

In two of the conditions the subjects had been informed that they would have an opportunity to judge the partner's work by means of the shocks after he had evaluated their performance. These people then presumably expected that they would have a socially sanctioned opportunity to retaliate for the harsh treatment they had received (Expect to Aggress group). The remaining men had been led to believe they would have no opportunity to administer shocks to their partner (Expect Not to Aggress group). About 5 minutes after they had received their anger arousing shocks all subjects were shown what was supposedly the partner's solution to another problem. The expectations were met in only half of the cases. Only half of the people in the Expect to Aggress condition were permitted to administer shocks at this time (the partner did not receive any shocks, of course), and only half of those in the Expect Not to Aggress condition were not able to shock their partner. The remaining people were told that the experimenter had made a mistake in reading the schedule. Thus, the other half of the men in the Expect to Aggress group found that they would not be able to give the shocks they had anticipated giving, and the other half of the subjects in the Expect Not to Aggress

---

aggressive behavior might bring disapproval, and then leaned over backwards to avoid showing such "dangerous" (for them) behavior. Rosenbaum and deCharms report these subjects exhibited very little open aggression when they had the opportunity to retaliate against their frustrater and, as they acknowledge, the "catharsis" found for this group was probably more apparent than real.

[9] The experiment was conducted by the writer and Ronald Luehrig.

treatment learned that they actually did have an opportunity to evaluate their partner via the electric shocks.

When all the subjects had looked at their partner's solution and had administered as many shocks to him as they thought desirable—that is, if these shocks had been permitted—they were given a brief questionnaire on which they rated, among other things, how much inner tension they felt at the time. Shortly after this questionnaire was completed, the subjects were again shown what was supposedly the partner's performance on yet another problem and this time all subjects had an opportunity to "rate" the performance by giving the partner shocks. A second questionnaire was then administered assessing the subjects' felt tension level on this occasion. The subjects also rated their partner on a brief adjective check list used as a hostility measure in other research in our project. The experiment was then concluded and the deceptions were explained.

In addition to the four conditions created by means of the experimental manipulations described up to this point, a nonangered control group was also established. These people had received the minimum one shock as the evaluation of their work, had expected to be able to shock their partner in return, and actually were permitted to administer the shocks to him. They also completed the various questionnaires at the appropriate times.

Our most important question concerns the effects of the subjects' initial expectations. Would those angered men who had anticipated being able to shock the anger instigator and who then were deprived of this aggressive opportunity display signs of stronger frustration than the other subjects before the end of the session? The reasoning employed in the present paper suggests that this should be the case.[10] In the words of the more behavioristically inclined psychologists, when we expect to obtain a certain satisfaction we made "fractional anticipatory goal responses." This phrase is a good one for our purposes. First of all, it suggests that the goal is symbolically present in our thoughts. Expecting to aggress means the aggression goal is symbolically represented. This representation serves as a cue evoking aggressive responses. We can say the thought of the goal, the anger instigator, is the cue eliciting implicit aggressive responses—if we are angry and primed. Thus, the phrase also implies there is an ongoing aggressive response sequence if only in the thoughts of the angered men in the Expect to Aggress condition. Those angered people not anticipating an opportunity to aggress against their tormenter should be less likely to have such an aggressive sequence in operation and, consequently, should be less bothered by not being permitted to attack him.

In the present experiment, then, the former group deprived of an expected opportunity should report a higher level of experienced tension before they are given the final chance to shock their partner. Further, when this opportunity is provided, they should exhibit stronger overt hostility; the frustrated aggressive sequence should heighten the arousal state leading to more intense aggressive acts at the end of the session.

The results with our shock measures are summarized in Table 2. The first

---

[10] Some might argue that any heightened predisposition to aggression stems from the thwarting of a "set" or (if they prefer other terminology) "expectancy." But does the use of such concepts result in a better explanation? I do not think so. What is a set? It may consist of an implicit but nevertheless ongoing chain of responses.

TABLE 2

MEAN NUMBER OF SHOCKS REPORTED RECEIVED AND NUMBER GIVEN

Measure	Angered				Non-angered
	Expect to shock		Expect not to shock		Expect to shock
	Able	Not able	Able	Not able	Able
1. Number shocks reported received by subject on first problem:	6.91	6.81	6.22	6.88	1.00
2. Number shocks given by those permitted to "rate" partner's first problem after arousal:	4.17	—	4.17	—	3.03*
3. Number shocks given as judgment of partner at end of session:	3.72_b	4.42_a	3.88_b	3.64_b	3.08_c**

Note.—Except for Measure No. 1, the shock data have been transformed using the $\sqrt{x} + \sqrt{x+1}$ transformation in order to facilitate statistical analysis. Each mean is based on 16 cases.

* The nonangered group gave significantly fewer shocks to the partner on the first problem than the combined anger groups ($F = 17.40$, $p < .01$).

** Differences among conditions were tested by t test employing the residual mean square from the preliminary analysis of variance in the error term. Cells having a subscript in common are not significantly different at the .05 level of confidence.

line of data indicates that the people in each angered condition reported receiving essentially the same number of shocks from their partner. That the excessive punishment was anger provoking is suggested by the condition differences in shocks given for the partner's first performance (line 2). The two presumably angered groups who were permitted to shock their partner at this time gave reliably more shocks to him (when combined) than the nonangered control group. Opportunity to attack the anger instigator in the adjoining room produced cues evoking aggressive responses from the angered men. Note, however, that there was no difference between the two provoked groups in number of shocks administered at this time; the Expect to Shock–Able to Shock men displayed the same level of overt aggression in their first "judgments" as the subjects who had not anticipated being able to retaliate. The information given the former people that they could shock their partner right away had apparently not "stirred them up" in the short time before aggression was permitted.

TABLE 3

MEAN RATING OF THE PARTNER ON THE ADJECTIVE CHECK LIST

Angered				Non-angered
Expect to shock		Expect not to shock		Expect to shock
Able	Not able	Able	Not able	Able
19.25_{ab}	22.44_a	17.44_{ab}	20.38_{ab}	15.06_b

Note.—The higher the score the greater the indicated unfriendliness. Each mean is based on 16 cases. Differences among conditions were tested by t test employing the residual mean square from the preliminary analysis of variance in the error term. Cells having a subscript in common are not significantly different at the .05 level of confidence.

The "expect to aggress" information evidently did start an implicit aggressive sequence in operation in the angered subjects, however. As the last line shows, when the anticipated

TABLE 4

MEAN SCORES ON TENSION INDEX

	Angered				Non-angered
	Expect to shock		Expect not to shock		Expect to shock
	Able	Not able	Able	Not able	Able
1. After seeing the partner's first performance:	20.48_{ab}	21.83_{a}	19.27_{bc}	20.38_{abc}	20.24_{bc}
2. After giving partner the final shocks at end of session:	20.44_{ab}	20.89_{ab}	18.82_{c}	21.13_{ab}	20.27_{abc}
3. Mean change:	−0.04	−0.94	−0.45	+0.75	+0.03
4. Number in group decreasing in reported tension:	9	11	6	6	7

Note.—The tension data were transformed using the $\sqrt{x} + \sqrt{x+1}$ transformation. The two sets of tension scores were subjected to one "repeated measures" analysis of variance with the t tests between condition means employing the residual mean square from this analysis in their error term. Cells having a subscript in common are not significantly different at the .05 level of confidence. Each mean is based on 16 cases. The higher the score the higher the reported tension.

aggressive activity was blocked in the Expect to Shock–Not Able condition, a heightened arousal resulted so that the men in this group administered more severe punishment to the anger instigator at the end of the session than did the people in any other group. The thwarting of the instigated aggressive sequence apparently had to persist for some minutes before the increased inclination to aggression could be revealed in stronger overt hostility.

But while there are indications of a frustration-strengthened arousal state, Table 2 contains no evidence of a cathartic reduction in aggressive tendencies following the initial display of hostility. The subjects in the Expect to Shock–Able group gave just as many shocks to their partner at the end of the session as the people who had neither expected nor had been able to shock their peer earlier.

As can be seen in Table 3, essentially the same pattern of results was obtained with the adjective check list measure of hostility administered after the subjects had given their partner the final shocks. In this case the Expect to Shock–Not Able group was more unfriendly to their peer than any of the other provoked conditions, although this difference did not attain statistical significance. These men, however, were the only angered men expressing reliably greater unfriendliness to the partner than the control group. Also, there is no evidence of a cathartic hostility reduction with this measure either.

But while the anger reduction version of the catharsis hypothesis receives little support, there is some evidence for the contention that aggressive behavior may lead to a lessening of experienced tension under some conditions. The subjects in the present experiment rated their mood on three separate scales on two separate occasions. The occasions were: (*a*) after they had been provoked by the partner and had been given or not given an opportunity to retaliate, and (*b*) after all subjects had administered shocks to the partner at the end of the session. The mood items were three bipolar scales, relaxed—tense, happy—sad, and angry—not angry, on which the subject had to indicate his feelings dex was established by adding the at the time. A combined tension index was established by adding the three scores, taking care to have a high score signify "high felt tension" in each case. Now, how did the conditions differ in these tension scores on the two occasions? The results are summarized in Table 4.

As we would expect, the people in the Expect to Shock–Not Able condition on Occasion 1 reported feeling more tense than the men in any of the other groups. While they did not differ reliably from the other angered conditions, these people were the only ones to have a mean tension score significantly higher than that in the nonangered control group. The thwarted ongoing aggressive sequence in the subjects who had anticipated being able to shock their peer but then were unable to do so apparently produced a relatively high level of felt tension within them.

On Occasion 2, after all subjects had been given their final socially sanctioned opportunity to aggress against their partner, the mean tension level in the Expect to Shock–Able condition decreased so that it no longer differed from that in the control group. There was not a significantly greater decrease in this group than that in any other condition, but the general tendency is certainly as we would have predicted: The people who had expected to attack him earlier but could not do so, and thus, whose implicit aggressive responses were prevented from reaching completion, tended to experience a fairly strong tension reduction when they finally were permitted to aggress against their tormenter.

That the tension decrease in this condition is more than a regression toward the mean is suggested by the data given in the last line of the table. Since several of the subjects in each of the groups, particularly those in the Expect to Shock–Able condition, showed unusual tension increases, comparisons were also made of the number of men in each group exhibiting a reduction in felt tension. The results of this analysis again indicate that the "expect to aggress" information had instigated an implicit aggressive sequence in the angered subjects which attained at least some degree of completion with the final shocks. Combining the two Expect to Shock groups and the two Expect Not to Shock conditions, we find a reliably greater proportion of the subjects in the former "treatment" decreased in reported tension from Occasion 1 to 2 ($\chi^2 = 4.00$, $p = .05$).

Completion through Symbolic Injury. We have already noted that an aggressive response sequence may attain completion through symbolic injury of the anger instigator. Hokanson and Burgess (1962), it will be recalled, had observed a reduction in physiological tension in previously angered subjects permitted to rate their tormenter on a questionnaire.[11] Additional indications of the value of symbolic injury are found in research from our Wisconsin laboratory. In an experiment by Berkowitz, Green, and Macaulay (1962), when the subjects were told their peer had done a bad job on the assigned task, the angered group not permitted to aggress against him indicated they felt reliably better than did the people in any other condition. The provoked people may have been pleased to learn the instigator had done so poorly. In other words, they may have believed that the experimenter had injured their frustrator symbolically by "showing him up." Some support is given this interpretation by the findings in a more recent experiment by the same researchers. Under similar conditions, angered subjects shown that their instigator had done a very poor job had a reliable increase in the "goodness" of their mood. Such a change did not occur in any of the nonangered groups or in the aroused people who were led to believe the instigator had done good work.

---

[11] In a later experiment, published as this paper went to press, Hokanson, Burgess, and Cohen (1963) found that frustrated college students given an opportunity to administer electric shocks exhibited a reduction in blood pressure level (in comparison to that displayed by frustrated people not permitted to aggress) when the object of their attack was the frustrater. There were lesser degrees of reduction in physiological tension, and not significantly different from that shown by the frustrated control group, when

SUMMARY

What can we conclude from these findings? Should we encourage people to express whatever hostile wishes they might have? An angered person does not necessarily make aggressive responses. Among other things, stimuli associated with the anger instigator must be present, whether in the objective situation or in his thoughts, if the hostile activity is actually to occur. By encouraging the angry person to attack his frustrater we may actually provide him with these aggression-evoking cues (as well as lower his inhibitions by justifying such hostile acts). However, if he is thinking about aggressing against his tormenter and wants to do so—in essence, symbolically giving himself hostility eliciting cues—an inability to attack him is a frustration and may heighten his already aroused aggressive predispositions. Should he then be able to aggress against his anger instigator he may feel better as his internal, frustration-engendered tension is reduced. "Feeling better," however, does not in itself mean there is a lessened likelihood of attacking the anger instigator again on some future occasion. The individual may have acquired an unfavorable attitude toward the anger instigator so that on some future occasion he may be quick to react aggressively to him when he is frustrated either by this person or someone else (cf. Berkowitz & Green, 1962).

REFERENCES

ADORNO, T., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D., & SANFORD, R. *The authoritarian personality.* New York: Harper, 1950.

ALLPORT, G. W. *The nature of prejudice.* Cambridge, Mass.: Addison-Wesley, 1954.

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

BERKOWITZ, L. *Aggression: A social psychological analysis.* New York: McGraw-Hill, 1962.

BERKOWITZ, L., CORWIN, R., & HEIRONIMUS, M. Film violence and subsequent aggressive tendencies. *Publ. Opin. Quart.*, 1963, **27**, 217–229.

BERKOWITZ, L., & GREEN, J. A. The stimulus qualities of the scapegoat. *J. abnorm. soc. Psychol.*, 1962, **64**, 293–301.

BERKOWITZ, L., GREEN, J. A., & MACAULAY, JACQUELINE R. Hostility catharsis as the reduction of emotional tension. *Psychiatry*, 1962, **25**, 23–31.

BERKOWITZ, L., & RAWLINGS, EDNA. Effects of film violence on inhibitions against subsequent aggression. *J. abnorm. soc. Psychol.*, 1963, **66**, 405–412.

---

the provoked subjects could shock only substitute objects such as the frustrater's assistant. Completion apparently is more satisfactorily achieved through attacks on the frustrater than through aggression against substitutes.

BROWN, J. W., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.,* 1951, **48**, 465–495.

BUSS, A. H. *The psychology of aggression.* New York: Wiley, 1961.

CAMPBELL, B. A., & SHEFFIELD, F. D. Relation of random activity to food deprivation. *J. comp. physiol. Psychol.,* 1953, **46**, 320–322.

DOLLARD, J., DOOB, L., MILLER, N., MOWRER, O., & SEARS, R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

HARTMANN, H., KRIS, E., & LOEWENSTEIN, R. M. Notes on the theory of aggression. In, *Psychoanalytic study of the child,* Vol. 3–4, New York: International Universities Press, 1949.

HESS, E. H. Ethology. In R. Brown, E. Galanter, E. H. Hess, and G. Mandler, *New directions in psychology.* New York: Holt, Rinehart & Winston, 1962.

HINDE, R. A. Energy models of motivation. In, *Symposia of the Society for Experimental Biology,* No. 14. Cambridge: Cambridge Univer. Press, 1960.

HOKANSON, J. E., & BURGESS, M. The effects of three types of aggression on vascular processes. *J. abnorm. soc. Psychol.,* 1962, **64**, 446–449.

HOKANSON, J. E., BURGESS, M., & COHEN, M. F. Effects of displaced aggression on systolic blood pressure. *J. abnorm. soc. Psychol.,* 1963, **67**, 214–218.

HOKANSON, J. E., & SHETLER, S. The effect of overt aggression on physiological tension level. *J. abnorm. soc. Psychol.,* 1961, **63**, 446–448.

KLUCKHOHN, C. Group tensions: Analysis of a case history. In L. Bryson, L. Finkelstein, & R. MacIver (Eds.), *Approaches to national unity.* New York: Harper, 1945. (Cited in G. Allport, *Nature of prejudice,* Cambridge, Mass.: Addison-Wesley, 1954.)

LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

MALMO, R. B. Activation. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology.* New York: Basic Books, 1962.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm soc. Psychol.,* 1948, **43**, 155–178.

MORGAN, C. T. Physiological theory of drive. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

ROSENBAUM, M. E., & DECHARMS, R. Direct and vicarious reduction of hostility. *J. abnorm. soc. Psychol.,* 1960, **60**, 105–111.

SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child-bearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.,* 1953, **47**, 135–234.

SHEFFIELD, F. D. A drive induction theory of reinforcement. Paper read at Psychology Colloquium, Brown University, November 1954.

THIBAUT, J. W., & COULES, J. The role of communication in the reduction of interpersonal hostility. *J. abnorm. soc. Psychol.,* 1952, **47**, 770–777.

TINBEGEN, N. *The study of instinct.* Oxford: Clarendon Press, 1951.

VON HOLST, E., & VON SAINT PAUL, URSULA. Electrically controlled behavior. *Scient. American,* 1962, **206**, 50–59.

WEATHERLEY, D. Maternal permissiveness toward aggression and subsequent TAT aggression. *J. abnorm. soc. Psychol.,* 1962, **65**, 1–5.

WILLIAMS, R. M., JR. *The reduction of intergroup tensions.* SSRC Bull. No. 57. New York: Social Science Research Council, 1947.

(Received February 4, 1963)

Psychological Review
1964, Vol. 71, No. 2, 123–140

FACTS, FACTORS, AND ARTIFACTS:

THE QUEST FOR PSYCHOLOGICAL MEANING

RICHARD W. COAN

University of Arizona

Issues relating to the psychological interpretation of factors are surveyed, and possible solutions are suggested. The most basic arguments regarding the reality and interpretability of factors are examined in terms of preferences for different modes of verbal formulation, preferences for different types of theoretical constructs, and confusion regarding the function of constructs. The multiplicity-parsimony issue largely dissolves when examined in the light of the functions of scientific theory, but the relevant concepts of simplicity and unity need more operational analysis. Arguments regarding causal status stem largely from widespread failure to analyze causal concepts and focus directly on the component issues. The closely related problems of descriptive generality are examined in the light of a modified hierarchical model.

During the past few decades, there has been increasing use of the various methods of factor analysis in many areas of psychology. Throughout this same period, controversies regarding the proper psychological interpretation of the products of these methods have persisted in virtually unmodified form. It is the contention of the present writer that many of the points of dispute have remained unresolved simply because of a failure of writers in most quarters to subject the underlying issues to sufficient critical analysis. It is the purpose of the present paper to examine these basic issues and offer some suggestions for possible solutions. Since arguments have arisen chiefly in the context of the traditional R-technique analysis of intellectual performance and personality, it is in this area that we shall focus primary attention. The basic issues that we shall consider, however, are broader than this realm of application, and the conclusions we shall draw should be considered subject to logical extension to other applications of factorial techniques.

THE THEORETICAL STATUS OF FACTORS

Perhaps the question of greatest priority is whether factors possess any psychological meaning at all. Allport (1937) once noted that many factors fail to make good psychological sense and that they thus run the risk of being considered mathematical artifacts. Others have argued that we should not attempt to make psychological sense of factors as such. Anastasi (1938), for example, regards the factor as a useful mathematical expression of correlational trends. As such, it is no more an "artifact" than the standard deviation, but it should not be construed as representing a psychological entity. Essentially the same position has been expressed in various publications by Thomson (1935, 1951). Other writers, like Guilford (1940, 1959a), have consistently argued that factors may represent a kind of psychological reality and that, if properly derived from properly designed research, they are not simply "statistical coefficients." Bartlett (1937), who allies himself fairly closely with Thomson, has even argued that Thomson's own theory provides a basis for attributing some degree of reality to factors.

These apparently divergent positions are not so contradictory as they might appear at first glance. It must be granted that, by whatever procedure it is derived, a factor remains a mathematical function of the correlations with which one begins. If these correlations express covariation among psychological variables, the factor will, in a somewhat different way, represent such covariation. Depending on the circumstances of observation, correlations may reflect peculiarities of a particular occasion of testing or of a particular sampling of subjects. A resulting factor is subject to the same influences, as well as those dependent on the overall composition of the test battery. The use we can make—practical or theoretical—of a given correlation depends on the circumstances under which it is replicable. The same thing is necessarily true for a factor. Whether the correlations are governed essentially by relatively transitory or by stable influences, they represent a kind of psychological "reality" which is inevitably mirrored in the factor. To the extent that a factor fails to be replicable, of course, it may be pointless to interpret it, since the underlying reality consists in large part of "chance"—i.e., of influences which we have failed to control or which are irrelevant to our immediate theoretical interests.

Since everyone who examines the issue must grant that factors possess psychological reality in the sense just considered, the truly important differences in viewpoint are those relating to the extent to which, and to the way in which, we may justifiably draw theoretical interpretations or inferences from the observed factor. Thomson's (1951) position is a highly conservative one, for he sees the value of factor analysis as being almost entirely practical, rather than theoretical. Most other prominent factor analysts, however, are interested in assigning psychological interpretations to factors for theoretical purposes. In examining the various viewpoints regarding psychological interpretation, we should bear in mind that writers in this field generally require the term *factor* to perform a double service. Some confusion arises from the fact that the word may refer either to the statistical entity yielded by our mathematical operations or to a psychological interpretation of this. There might be some value in using different terms for the statistical concept and the psychological (or more generally, theoretical) concept. Cattell (1957, 1961) has apparently recognized this need in employing *source trait* to represent his psychological interpretation. In a recent paper (Cattell, 1961) he has underscored the fact that the factor as a statistical entity relates in a variable and imprecise way to the underlying source trait. He has not seen fit to remedy the terminological ambiguity, however, for he continues to employ the word factor in a double sense. It is doubtful that any terminological substitution proposed now would gain universal acceptance. In some discussions, however, the introduction of subscripts (e.g., using $factor_s$ and $factor_p$ to represent the statistical and psychological concepts, respectively) might make for clearer communication.

Some of the confusion stemming from the terminological ambiguity is apparent in the recurrent discussions of such questions as whether factors are genotypic or phenotypic. As a statistical entity, the factor is strictly a function of covariation currently present in the observed variables and must be regarded as representing a phenotypic correlational trend. The phenotypic condition presumably has a history and corresponds in some way to genotypic determiners. The essential question is whether from the statistical entity we can make a fairly direct inference to a more or less unitary genotypic influence. If so, the factor may be interpreted in terms of this influence and considered, as a psychological concept, to be genotypic. Nothing in the nature of the factor procedures could guarantee this consequence. Apart from a logical problem regarding the nature of "unity" (which we shall consider below), the question is an empirical one, whose answer will probably vary depending on the character of the variables we choose to measure. In any case, a "factor" in one sense *must* be phenotypic; in another sense, it *may* be genotypic.

The various positions seen with regard to the psychological interpretation of factors may be thought of in terms of preferences for various kinds of theoretical concepts. Elsewhere, the writer (Coan, 1963) has distinguished between *adspective* concepts, which are definable in terms of observed entities, events, or relationships, and *ultraspective* concepts, whose definitions contain factual content that exceeds the bounds of mere induction from past observation. Interpretation of a factor in terms of an adspective concept might consist of avoiding any inference beyond a present set of loadings and treating the factor simply as a function of the observed variables. In such a case, one might say that interpretation is actually avoided, since it is allowed to rest with the product of our statistical labors. An ultraspective concept may take basically either of two forms—a hypothetical concept or a fictional concept. The hypothetical concept refers to entities, events, or relationships which have not been observed but which are regarded as potentially observable. The fictional concept, on the other hand, refers to entities, events, or relationships which are not assumed to correspond to fact and which are thus not regarded as potentially observable.

A factor interpretation ordinarily takes the form of a hypothetical concept when the factor is interpreted in terms of a particular determinant lying outside the range of observation of the immediate study. This determinant might be a biological structure or condition, a form of social experience, etc. Even if the interpretive concept corresponds essentially to previously observed entities and events, in such cases it will usually embody undemonstrated functional relationships as well. A more common form of interpretation consists of equating the factor with a general behavioral characteristic—say, an ability or a personality trait—that might be defined operationally in terms of the kinds of performances that have been observed. This sort of interpretation ranges between the adspective and the hypothetical. It is clearly adspective in the rare instances where the tested performances are regarded as exhausting the realm of behavior in terms of which the concept is defined. It is still adspective, though less clearly so, when the observed behaviors are regarded as representatives of classes to which they obviously belong, so that factor interpretation generalizes, in effect, to similar batteries, and to occasions other than that on which observations have been recorded. The induction, of course, may or may not be supported by subsequent observations. The interpretation may be considered to take the form of a hypothetical concept when the alleged general behavioral characteristic involves a clear qualitative departure from the observed performances—when, for example, the sampled behaviors are treated as indirect expressions of a more fundamental trait that is truly central to the factor. Even here, the interpretive concept preserves the basic features of the adspective concept in so far as it is definable in terms of behaviors that have been observed at some time. Its hypothetical character resides in the implication of a pattern of covariation that goes beyond what has been demonstrated.

Some writers in this field object to any identification of a factor with an ability or trait, apparently regarding such interpretation as a form of reification or hypostatization. Anastasi (1938), for example, criticized Thurstone's treatment of factors in this vein, accusing him of resurrecting faculty psychology by introducing a sharp dichotomy between ability and behavior. Thurstone's factor labels, however, seem largely to constitute abstract disposition concepts operationally definable in terms of behavior. In formulating such abstractions, Thurstone clearly reveals a preference for generalizing further beyond the immediate data than Anastasi would be inclined to do, but this hardly justifies a charge of hypostatization.

It is possible that the core of the early dispute between Anastasi and Thurstone lies simply in a difference regarding preferred modes of verbal formulation. But it must be granted that many theorists do fashion interpretations in such a way as to imply reference to tangible underlying entities, and Anastasi is not alone in her concern with this problem. Burt (1941), too, has expressed alarm at what he perceives as a widespread tendency to reify factors. Thurstone leaves himself open to this sort of criticism to the extent that he speaks of factor analysis as a method for isolating "psychological entities" or "mental faculties" (cf. Thurstone, 1935, 1938). While he regards his faculties as being fundamental in the sense that they possess greater cohesiveness and stability than do many specific forms of behavior, however, it is nowhere evident that he regards them as having an independent existence apart from the behavior in which they are expressed. The situation is apparently comparable for most of the other writers criticized by Burt.

Of the major factor theorists, Spearman (1927) appears to come closest to an interpretation of factors in terms of entities enjoying a concrete status not reducible to behavioral expression. He argues effectively against equating his *g* factor with any particular kind of mental operation, showing by his data that it appears as a factor in different kinds of operations. He argues instead for an interpretation of *g* in terms of the "amount of general mental energy," while he interprets his *s* factors in terms of the "efficiency of specific mental engines." He speculates on the possibility of interpreting the "engines" more specifically as different neural systems, but his primary interest is in factor *g*, for which available information affords little guidance for formulation of specific physiological hypotheses. For his purposes, he is content to leave his "mental energy" in the form of a formalized and unspecified hypothetical concept. For that matter, he feels that the concept would effectively serve the purpose of bringing systematic theoretical order into the realm of intellectual performance even if it were denied any definite physical reality whatever. For those who insist on such a denial, Spearman is willing to say that mental phenomena function *as if* such an energy existed. This concession would accord *mental energy* the status of a fictional concept. Psychologists are more generally inclined today to see heuristic value in fictional physical models than they were at the time of Spearman's writing, but the fictional component of the *energy* concept may be too vague to serve any effective purpose. Simply as a construct, of course, the concept serves the basic purpose for which Spearman intended it.

Other writers, such as Thurstone (1938, 1948) and Guilford (1940), indulge in comparable physiological speculation. For both Thurstone and Guilford, however, it is clear that a factor as psychologically interpreted carries no necessary implication of biological status. Instead, the physiological basis underlying the factor is treated as a matter of additional hypothesis. Thus, Guilford suggests that some factors will be found to have well-defined biological correlates in the individual, while others will not. In a given factor the physiological basis might be localized to a specific tissue or it might reside in some more general property of the organism. Thurstone notes a close parallel between his first- and second-order factors and Spearman's *s* and *g* and suggests, in accordance with Spearman's formulations, that the primary factors may have a relatively localized basis, whereas the more general second-order factors are likely to relate to more central or general biological parameters.

A somewhat different trend in factor interpretation is that dominated by Burt (1941, 1949), who speaks of factors as "principles of classification." He objects to the use of terminology that appears to identify factors with concrete causes, whether in terms of "faculties," "entities," "abilities," "traits," or "energy." He insists that the factor does not represent a component of the individual. Rather it is an "abstract component" that provides a convenient frame of reference for description. Essentially the same view is reiterated by many other writers, including Moursey (1952), Vernon (1950), and Eysenck (1947, 1951).

The basic position expounded by Burt is probably accepted today by the vast majority of factor analysts. It is to Burt's credit that he has so clearly and competently brought the issue into focus. The writer is inclined to believe, however, that much of what Burt perceives as a differing theoretical orientation is simply a preference for a different mode of description. One may speak of faculties, primary abilities, and unitary traits without attributing to these the fundamental status of anatomical organs. Regardless of the descriptive labels which Thurstone chooses to employ, for example, there can be no question but that his basic argument for the use of factor methods is a defense of the scientific search for useful theoretical constructs, not of a search for absolute units (cf. Thurstone, 1947).

The issue is clouded, of course, because of the subtle nature of changes in psychological theory. Psychologists do not merely reclothe old concepts in new guises. They sometimes subject old terminology to reconceptualization befitting a change in outlook. Thus our cellars are stocked with both "old wine in new bottles" and "new wine in old bottles." In any case, factorists continue to identify factors with attributes of individuals—traits, abilities, etc.—but they seem generally to regard these attributes as rather arbitrarily chosen concepts justified by the needs of scientific theory construction. A clear methodological difficulty would appear to arise only when a researcher regards his interpreted factors as *the* ultimate objects of scientific investigation, as fundamental discoveries rather than theoretical constructions. His concepts then present a barrier to scientific progress, but not because they are hypothetical or constructive—for hypotheses and constructs are essential to scientific progress. They fail to serve science because they contain a hypothetical element that is not subject to confirmation. A factor interpretation in the form of a hypothetical concept is useful if the entities or relationships to which it refers can be examined in the light of further research. Similarly, a fictional concept may be helpful, since its ultraspective component is recognized as a convenient construction that is subject to replacement.

In the interest of scientific progress, the writer would argue for a free and flexible use of different kinds of interpretive concepts. Liberal speculation can best serve its purpose, however, if it is tempered with a clear recognition of the character and function of the concepts. An understanding of the nature of theory construction should obviate unnecessary delay in the abandonment of concepts that have outlived their usefulness.

It may be of interest to note that beyond the realm of factor theory, the issue of hypostatization appears to concern theorists recurrently. It probably arouses less alarm generally in psychology today than it did in the 1930s and 1940s. Whether there is an actual danger in construing noun labels as representing "real entities" or "immaterial substances," of course, depends on the meaning one attaches to these "real entities"—whether one regards them as having some kind of permanent or ultimate factual (truth) status equivalent to that of material objects. With their increasing sophistication with respect to the somewhat arbitrary nature of constructs as ingredients of theory, psychologists have become justifiably less preoccupied with this issue. Perhaps at the same time, by virtue of this sophistication, hypostatization has come to pose a less serious threat.

MULTIPLICITY VERSUS PARSIMONY

Another issue that has received considerable attention is that of the number of factors or determinants required, in theory, to account for the observed correlations. In the relevant literature, this issue has been thoroughly entangled with those in the section above. Much of the discussion has centered around a position that Thomson (1935, 1951) has called sampling theory. Thomson sees practical value in the products of factor analysis, since men do perform intellectual tasks *as if* only a few factors were operative, but he maintains that for scientific purposes we must think of mental performance as being governed by a vast multiplicity of determinants. Psychologically, these determinants might correspond to Thorndike's "bonds"; physically, they might correspond to neurones or neural arcs. Thomson has shown statistically that a large number of specific determinants may yield a single factor or a small number of group factors (though the relevance of this demonstration to psychological data has been questioned by other writers on various grounds).

A similar position has been championed by Tryon (1935, 1958), who presents a logical analysis of test performance in terms of concepts and adjustments as psychological determinants and genes as biological determinants. Tryon admits that external agencies, in the form of cultural or educational influences, may function in such a way as to create correlational clustering of psychological components, but he insists that in this case the broader unity reflected in the resulting factor resides in the external agency and not within the individual affected by it. Tryon objects to what he considers an improper emphasis on parsimony on the part of most factor analysts. He feels that if one is seeking psychobiological causes, it is not appropriate to seek a parsimonious set of factors, since the actual causes are numerous. Burt (1941) takes a similar stand on the matter of parsimony. He believes that simplicity of formulation might furnish some guarantee of truth in the physical sciences but that it cannot do so in a science like psychology, where the subject matter is much more complex.

Both Burt and Tryon make the questionable assumption that those who advocate parsimonious theory believe nature itself to possess the simplicity embodied in theory. Yet it seems plausible that most factorists today would agree essentially with the views expressed on this matter by Thurstone (1947) and Kelley (1935). Both of these men argue, in effect, for a group-factor position on the grounds that it provides an economical set of constructs. These constructs are viewed not as part of an underlying nature but as a scheme imposed on nature by human beings to serve human purposes. Both writers recognize that the greater the number of constructs or components adopted, the more accurately and completely we can represent the natural state of affairs, but they would also argue that scientific theory always simplifies and that the degree of simplification we introduce must be governed by our needs and capacities as human beings to comprehend.

The present issue is, of course, bound up with the ancient problem of atomism versus holism, even though Thomson, for one, insists that he is not an atomist. The advocates of multiplicity, in any case, seem generally to create a rather artificial dichotomy between factual description and generalized theoretical statement. They assume that there is a certain level of description—that of genes, neurone arcs, concepts, mental bonds—at which our statements will correspond uniquely to the "truth." This is presumably the level at which "things" are really "things"—but are they? If we were to probe further, might we not decide that each gene is but a vast hodgepodge of atomic particles, that each neurone arc contains a complex string of semi-independent components, and that each emerging concept is analyzable into a host of momentary awareness? It seems to be increasingly recognized by scientists today that even the statements that appear most objectively descriptive represent a good deal of abstraction and construction on the part of those who issue them. In this, they differ only quantitatively from highly generalized theoretical formulations, for scientific statements at any level are a product of human perceivers and thinkers, and they are created to meet a human striving for understanding.

From this reasoning, it would certainly follow that anyone has a perfect right to seek an understanding of human behavior and experience on the level of neurones, genes, or specific mental bonds. These components have the same sort of reality as the elements of which they are composed and the aggregates which they form, but they have no unique claim to reality. Let us grant that much valuable knowledge may be gained by studying anatomy with a microscope, but let us not deny the existence of hands and faces on the grounds that extensive research with this instrument does not reveal them. It is true, as Tryon claims, that broad mental factors represent somewhat arbitrary selections, but then any unit of analysis represents an arbitrary choice. The point is merely a bit more obvious in the case of a dimension than in the case of an entity. Mental factors may be just as meaningful as such arbitrary components of physique as height and weight. The fact that genetic analysis does not reveal them is beside the point.

In the case of both entities and attribute dimensions, the scientist must choose those units or components which are optimally useful for a given theoretical purpose. Thus, the cell is a more useful unit of analysis for most physiological problems than, say, the chromosome or gene. Similarly, the most convenient unit for many problems of physiological psychology may prove to be a fairly complex neural network rather than the individual neurone. It is appropriate to formulate scientific statements in terms of factors, in preference to bonds or genes, when our immediate theoretical interests lie at the level of description to which factors correspond. The writer would contend that the theoretical concerns of psychologists most often—but not always—lie approximately at the factor level.

One point on which Tryon, Burt, Kelley, and Thurstone would all apparently agree is that the law of parsimony is a man-made principle rather than a law of nature. They would agree that its purpose is convenience of description, while differing with regard to the scientific use to be made of such description. If we grant the general point, we must also grant, to be consistent, that no particular form of simplicity has any special claim to truth, and we must concede to Thurstone's critics the fact that Thurstone has not always been this consistent.

In the realm of factor methods, simplicity is subject to a variety of interpretations. The most basic one, on which factor analysts generally agree, is an economy of constructs. We seek to account for a given number of variables in terms of a smaller number of factors. Once we have agreed on the number of factors to be employed, we note that from a mathematical standpoint there is an infinite number of sets of factors that will account equally well for our data. Of these innumerable sets, Thurstone would argue that the one best set is that which best satisfies his criteria for "simple structure." Unfortunately, Thurstone (1938) has sometimes spoken as if his criteria for simple structure constituted some kind of fundamental discovery. In the writer's opinion, they must be considered an arbitrary set of rules, though if the individual investigator is permitted to use them elastically, they furnish the most generally satisfactory basis for choosing a rotational solution that anyone has yet devised.

It is widely recognized that the criteria for simple structure can be met more readily if obliqueness among rotated factors is permitted, but obliqueness itself constitutes an additional form of complexity. The writer (Coan, 1959) has previously argued that, on the whole, oblique solutions are superior to orthogonal ones, but it must be emphasized that the two types of solutions furnish incommensurable forms of simplicity —factor simplicity and model simplicity. In such a choice between construct models, neither alternative is inherently more valid than the other. For that matter, one might even argue, in line with Burt's (1941) reasoning, that an unrotated least-squares solution has its own special claim to simplicity. It is this solution that enables us to account maximally for the observed variance with any possible number out of a larger number of extractable factors. For purposes of theory construction, however, the unrotated solution is difficult to defend. It is too dependent on the accidents of variable selection to yield the necessary reproducibility from one study to another.

The folly in attributing ultimate truth to any one type of solution is underscored by the fact that our earlier choice of number of factors is inevitably arbitrary in any empirical problem. This decision depends primarily on the importance that we attach, respectively, to precision and simplicity of description, but it is complicated by the obvious fact that simplicity with respect to number of factors is inversely related to the attainable simplicity of individual factors in a rotated solution. In any case, this decision delimits the possible redistributions of variance through subsequent rotation. In the writer's opinion, we can usually best meet our theoretical needs by employing a liberal number of factors thoroughly rotated to an oblique solution.

In the writings of Thurstone and his followers, the desiderata of factor constructs are often spoken of in terms of "functional unity." To some audiences, to speak of a factor as representing a functional unity is to imply that it represents some underlying indivisible element. Thurstone, however, has long made it clear that the factors or "faculties" he seeks do not constitute ultimates and that they may be broken down into further elements (cf. Thurstone, 1938). In this, he agrees with those who regard factors as "principles of classification" subject to any desired degree of subdivision (cf. Burt, 1949; Moursey, 1952). In essence, to say that a factor represents a functional unity is simply to say that it represents a reproducible pattern of covariation.

The matter needs to be examined further, however, because the concept of functional unity is often intended to imply some ill-defined cohesiveness of variables underlying this covariation. We must bear in mind that R-technique correlations and the factors that we derive from them represent directly correlational trends running *through a group* of subjects. It is important that we entertain the question of the extent to which we can properly infer from this covariation a functional interdependence of variables *within individuals*. In proposing a distinction between *object factors* and *population factors*, the writer (Coan, 1959) has previously suggested that the inference will prove to be justified in some cases but not in others. The issue is related to that raised long ago by Allport (1937), who said that factors represent only "average tendencies." In speaking of the "statistical" establishment of traits, however, even Allport (1931) appears to assume that the kind of inference involved here is warranted, at least up to a point. Moderate support for the inference has been furnished in recent years by the partial correspondence of R-technique and P-technique factors derived from a common measurement realm. Be that as it may, the question is still with us, and it deserves closer attention if we are going to build psychological theories on a correlational foundation. On logical grounds, we can at best argue for the plausibility of an inference of dynamic interdependence. We obviously need more in the way of direct empirical investigation of the matter.

In future explorations of this problem, we should bear in mind that apart from its fundamental meaning in terms of transpersonal covariation, *functional unity* may imply any of several different kinds of interrelationships. For one thing, it may imply a high level of causal interaction among elements, each directly affecting the others to a greater extent than it affects elements forming other factors. A second possibility is that constituent elements of the functionally unitary trend have a common source in a unified maturational process or a unified determining environmental context. Certain learned performances (e.g., that displayed in a vocabulary test) clearly illustrate this kind of unity. A third possibility is that the constituent elements display a common behavioral effect, or at least a convergence of effect, which tends to reduce their mensural separability. Whatever the multiplicity of influences, anatomical organs and physical dimensions display this sort of unity. Obviously some psychological or behavioral factors may function in the same way. These three conditions are neither mutually exclusive nor necessarily concomitant. It is almost inevitable that, in forming psychological interpretations of factors, we shall assume some kind of functional unity beyond that directly demonstrated by group correlation data. If we are to make the most precisely meaningful interpretations and formulate the most appropriate hypotheses, we must remember that "functional unity" is not itself a unitary concept.

THE CAUSAL NATURE OF FACTORS

Those who object to the "reification" of factors often express concern over the possibility that factors will be interpreted as causal entities. Thus, warnings against causal interpretation have been uttered by Thomson (1935), Anastasi (1938), Eysenck (1947), and Burt (1941, 1949). Burt (1941) does grant that factor analysis may yield products corresponding to causal entities, though the process of factorization does not guarantee this. Its immediate products, in any case, are phenotypes rather than genotypes. Burt further argues that explanation in terms of causal agencies is, after all, an outmoded approach which has no place in twentieth century science.

The counter view has long been advocated by Cattell (1957), whose interpretation of the factor as a *source trait* is deliberately intended to imply that the factor operates as a cause with respect to the observed "surface" correlations. Even in those occasional places where he makes a distinction between factor and source traits and speaks of the source trait as the "genotype of which the factor is the phenotype," there is still implicit the notion that the factor corresponds rather closely to a cause, in any case more closely than the correlations from which it is mathematically derived. Cattell assumes that the cause may take any of a variety of possible concrete forms. It may be something physical, physiological, social, or psychological in nature. Though the issues involved in the controversy between Burt and Cattell are basically more logical than factual, Cattell has recently sought to provide empirical support for his position (Cattell & Sullivan, 1962).

It has already been suggested that one source of difficulty is the elastic usage of the word *factor*. A more serious problem in the present controversy, however, is the failure of those who take part in it to agree on a common concept of causality. The word *cause* has long been used in a variety of ways, and despite a perennial concern with the matter extending far back into antiquity, no standardized usage has ever been adopted in philosophy or science. The concept displays two general properties which are fairly constant. First, it refers to something which is somehow functionally related to an effect. Second, whether we speak of *a* cause or of *the* cause, we perform an arbitrary selection of some ingredient or aspect out of a more inclusive determining field of interaction. However the cause is designated, it constitutes but one of innumerable possible abstractions that might be treated similarly.

A causal concept may embody any of several possible kinds of functional relationships. Thus, a "cause" may constitute any of the following: (*a*) an antecedent event (fire or the act of igniting as a cause of smoke), (*b*) a disposition property defined in terms of, and manifested in, the effect (gravity as a cause of falling, the magnetic property of an iron bar as a cause of the movement of iron filings), (*c*) a whole of which the effect is a part or a specific case (A's greater general intelligence as a cause of his solving Problem X more rapidly than B does), (*d*) a part of the whole manifesting the effect (helium as a cause of the dirigible's rising), (*e*) a characteristic of the system in which the effect occurs.

These five categories are neither exhaustive nor mutually exclusive. The first refers to causes in a genetic or historical sense. The antecedent event may be one which precedes the effect either invariably or with sufficient reliability to permit better than chance prediction from cause to effect or from effect to cause. The remaining categories represent ahistorical interpretations of cause. Of these, the last is the most general, for it obviously overlaps with *b, c,* and *d*. Category *b* offers the most clearly redundant form of explanation, except in so far as the cause embraces a greater variety of events than the effect explained—in which case the cause is also of Type *c*. In a broad sense, perhaps all causal explanation is ultimately redundant.

Those who speak of factors as causes generally do so in a broad ahistorical sense, adopting most clearly Interpretation *e* but making varying use of the other interpretations. Thus the factor, as a basic characteristic of the behaving organism, is regarded as a constituent of the contemporary field in which a given behavioral effect occurs (*e*). At the same time, it may be defined in terms of a predisposition to the behavioral manifestations constituting its effects (*b*), but viewed as possessing greater generality than any given effect so that it is subject to various forms of expression (*c*). While not defined in terms of antecedents, it is commonly assumed to be traceable to earlier biological or experiential sources and to have a longer prior history than does its immediate effect. It is this temporal generality, or stability, of the factor that makes the factor concept a tool for prediction as well as for generalized description.

The notion of generality, both in a temporal sense and with respect to diversity of effect, is evidently basic to Cattell's concept of cause. After a careful logical analysis of the relationships among loaded variables, primary factors, and second-order factors, Cattell (1957) concludes that specific variables must be regarded as consequences of factors and factors as causes of variables. For the same reason, second-order factors must be regarded as playing a causal role in relation to primary factors.[1] These conclusions follow from the facts that Cattell considers only if one adopts the premise that effects are more varied than underlying influences.

---

[1] It should be noted that the distinction between first- and second-order factors is an arbitrary ingredient of the factor model. As Schmid and Leiman (1957) among others have shown, the second-order factor is subject to mathematically equivalent representation as a broad factor at the first-order level. To apply his reasoning con-

Rather similar reasoning appears in a more recent work by Guilford (1959a), who tends to avoid causal terminology. Adopting a hierarchical conception of behavioral organization, Guilford says that "determination" within the hierarchy tends to operate in a downward direction, i.e., from the more general to the more specific characteristics. He believes also that prediction works most efficiently in the same direction.

Those who display greatest alarm at references to causes seem concerned lest we postulate some invisible motor within the machine. Yet basically, the arguments in favor of factors as causes rest on the idea that general characteristics in some way determine specific acts. A "cause" in this sense is a legitimate construct or abstraction as long as we make our usage clear. It is justifiable to speak of the whole as determining the part in so far as we find it convenient to order our thinking in this way, but we must not pretend that we are stating a fundamental law of nature. If this ordering leads to better prediction than its opposite, it is largely because deduction is more efficient than induction. Undoubtedly, many psychological problems relating to "underlying influences" and "depth," both inside and outside the realm of factor theory, are operationally reducible to problems of generality versus specificity. No doubt, some confusion might be prevented if we avoided the use of such terms as *cause, depth,* and *genotype* in referring to general characteristics. The use of the word *cause* in respect to factors has led to a rather sterile controversy.

FACTOR GENERALITY AND LEVELS OF DESCRIPTION

There are a number of other problems that relate to the generality of factors and description. It is an obvious fact that factors constitute a more general descriptive level than the variables from which they are derived, and it is equally obvious that within the realm of factors and within the realm of experimental variables there is a range of variation in generality. The possible relationship between these descriptive levels and levels of organization in the mind or behavioral system has long occupied the interests of various theorists.

Most treatments of this matter are outgrowths of the hierarchical viewpoint formulated by Burt (1941, 1949). Burt himself traces his view of mental organization to Herbert Spencer. Burt distinguishes four types of characteristics corresponding to the four possible types of factors commonly called general, group, specific, and error. He regards the four-factor theorem as fundamental, in as much as an adequate conception of the mind requires recognition of characteristics on all levels from the universal to the accidental.

A number of other writers—including Vernon (1950), Eysenck (1951), Moursey (1952), and Guilford (1959a, 1961)—have adopted Burt's model. These men all accept a scheme involving four levels, though they vary somewhat in the psychological identities attributed to the levels. A system of four levels represents a rather clean set of logical distinctions when one is distinguishing types of factors or possible degrees of distribution of char-

---

sistently to the empirical realm, Cattell would have to argue that under some conditions a broad factor will bear a casual relationship to a narrower, linearly independent but cooperative factor of the same order.

acteristics. As we move from types of factors to corresponding levels of mental or behavioral organization, however, adherence to a fourfold scheme becomes much more clearly arbitrary. This should be obvious from the mere fact that a factor may change from one level to another depending on the test battery. Of course, such facts are generally recognized. Though Moursey attempts to make an empirical test of the hierarchical model, he notes the artificial simplicity of the model and insists that only an approximate correspondence between the model and actual mental organization is to be expected. Guilford apparently makes similar assumptions. Eysenck stresses the formal, abstract character of the model and the relativity of distinctions in factor level.

Both Eysenck and Guilford are primarily concerned with personality structure, and they both recognize essentially the same levels of behavioral organization—those of specific responses, habitual responses (in Guilford's terms, the hexis level), traits, and types. Both men recognize, too, that obtained factors do not bear any fixed relationship to these four levels. It should be granted that the sort of pyramidal hierarchy espoused by these writers affords a convenient framework in terms of which we can relate aspects of personality differing in generality of effect. For such a purpose, the four levels constitute justifiable constructs. A basic problem for which there is still no completely adequate solution, however, is that of deciding how to make appropriate inferences from empirical factors to these levels of behavioral organization. Early fixation on a fourfold scheme of stratification might well serve to obscure some of the underlying issues and interfere with progress in their solution. Perhaps we ought to give more explicit recognition to a basic continuum of generality and speak of degree of generality rather than level of organization. We shall then be better prepared to cope with a problem that involves, on the one hand, a variable spread of generality among factors, among experimental variables, and between factors and variables, and on the other hand, a fluid correspondence of these factors and variables to behavioral structures.

To see more clearly the nature of the problem, let us first distinguish clearly between *contextual generality* and *referent generality*. As used here, contextual generality refers to the extent to which a factor loads the experimental variables in a given set of data—i.e., it refers to generality within a given mathematical context, without regard to the psychological realm represented by the data. Thus in a given study, a so-called general factor would possess greater contextual generality than a group factor, and a second-order factor would ordinarily possess more than a first-order factor. For factors of all orders, contextual generality may best be expressed mathematically in terms of the total variance accounted for within the experimental variables. Referent generality, on the other hand, would refer to the scope of reference of the variable or factor in the behavioral or mental system in terms of which it is psychologically interpreted—i.e., the variety of behaviors or mental activities to which it relates and the degree to which it relates to them.

There is an imprecise correspondence between contextual generality and referent generality. Within a given set of factors, a second-order factor will probably possess greater referent generality than a primary factor, but a second-order factor in one study may possess less referent generality than a primary factor in another study. Furthermore, two first-order factors in a given set, with comparable loading patterns, may differ greatly in referent generality, depending on the variables loaded. The experimental variables themselves have equal mathematical status in a given study and are, hence, regarded as identical in contextual generality, but they may vary enormously in referent generality.

Among the conditions relevant to the two kinds of generality, we must recognize all of the following:

1. Selection of subjects: As Vernon (1950) has noted, both the degree and type of selection may affect the factor pattern. If a factor remains recognizably the same as we go from a relatively unrestricted sample to a sample restricted on relevant variables, the loadings decrease in magnitude. In short, its contextual generality tends to decline, though referent generality may remain unchanged. Vernon argues that a high degree of selection favors the emergence of group factors in place of a general factor.

2. Number of factors extracted from a given set of variables: Successive centroid and principal-axis factors contain progressively smaller variance, and subsequent rotation tends to distribute the available variance more evenly. Hence, each rotated factor within a large set will tend to be of smaller variance and possess both less contextual generality and less referent generality than a factor within a smaller set derived from the same variables.

3. Kind of rotational solution: Oblique rotation tends to redistribute variance more evenly than does orthogonal rotation. Hence, orthogonal factors tend to display a greater range of contextual generality, and correspondingly of referent generality, some being more general and some less general than any of the oblique factors. This trend is illustrated by a previous study of the writer (Coan, 1959).

4. Referent generality of scoring composite: The factor, of course, will display higher referent generality than the scored variables from which it is derived (unless it is a "singlet" factor), but the higher the referent generality of the variables, the higher will be that of the factor, whatever its contextual generality. The discrepancy between the factor and the variable in degree of referent generality will, of course, be determined by the other conditions treated in this list. The referent generality of the scored variable will be a function both of the basic measurement operation (rating of general traits, recording of specific movements, etc.) and of the mode of combination of scoring units (items) into the composite that constitutes a correlatable score.

5. Density of variable sampling: Vernon (1950) has noted that one may increase the contextual generality of a factor substantially by specializing the battery. In so far as the psychological identity of the factor is not altered in the process, referent generality would remain unchanged. Density of sampling, however, also determines what factors will emerge. We may expect relatively dense sampling within a given behavioral domain to yield factors not too far above the variables in degree of referent generality. Sparse sampling is conducive to the direct emergence of factors that would be of a higher order if dense sampling had been used.

A consideration of the above points suggests that, in place of a four-level hierarchy, we might better conceptualize the psychological implications of factors in terms of the scheme shown in Figure 1. In this figure, the vertical dimension represents a continuum

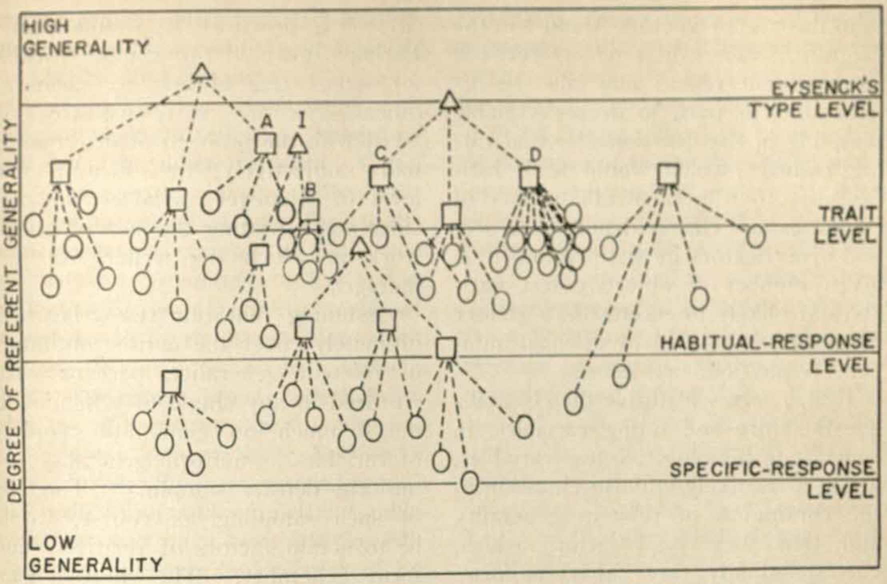


FIG. 1. A modified hierarchical representation of mental and behavioral organization.

of referent generality. The circles represent experimental variables, and the squares and triangles represent, respectively, first- and second-order factors derived from them. The elements thus shown in the figure may be construed as compiled from a variety of studies, the dotted lines indicating demonstrated loadings. Horizontal spacing may be regarded as an inverse function of correlational affinity (and perhaps of a corresponding relatedness within the individual), though if we wished to fit actual empirical data to such a figure, we should have to regard any horizontal array as the trace of a multidimensional configuration in which proximity faithfully represented correlation (by any convenient geometric convention).

As the figure is intended to suggest, variables, primary factors, and second-order factors may occur at all elevations of referent generality. Within a single study, their vertical succession is essentially orderly, but vertical discrepancies within the context of a single study are subject to variation. Factors B and C illustrate the effects of dense and sparse sampling of variables, while the difference between B and D is of the sort that might be governed by decisions regarding the number of factors to extract or the choice of rotational solution.

It is well known that different investigators in the personality field select variables of different degrees of referent generality and extract factors of different degrees of generality. The factors of Eysenck, for example, are obviously of higher referent generality than those of Cattell—largely because he extracts far fewer factors, but to some extent because of differences in method of rotation and variable selection. Within Cattell's work, there is probably considerable generality range in variable selection. Most notably, there seems to be a tendency for his objective test variables to be of higher referent generality than his rating and questionnaire variables, with the result that primary objective test factors tend to align with second-order rating and questionnaire factors, as in the case of Factors A and I in the diagram. The discrepancy between measurement realms may also be attributable, in part, to denser variable sampling in the questionnaire and rating realms (which would lead, incidentally, to a greater obliqueness of primaries and clearer definition of second-order factors in these realms). A given number of objective test variables are likely to encompass a greater range of behaviors than an equal number of questionnaire items.

It also seems plausible that Cattell's questionnaire and rating variables, in contrast to his objective test variables, are at a relatively uniform elevation in the continuum of referent generality and that they yield factors which, correspondingly, are fairly uniform. The objective-test variables undoubtedly cover a wide range of referent generality, and the resulting factors probably display a comparable vertical spread. Thus, Factor U. I. (T) 24 (Anxiety), which matches a second-order questionnaire factor, displays high loadings for a set of rather generalized verbal measures (cf. Cattell, 1957). In contrast, Factor U. I. (T) 22 (Corticalertia) loads a variety of simple tasks involving alertness and speed. It seems likely that this factor is of lower referent generality than U. I. (T) 24, unless we can assume that it draws from a sufficiently wide range of variables so that it is of considerably higher referent generality than the variables that it loads.

As Guilford (1959a, 1961) has pointed out, there is a danger of confusion in studies involving a mixture of variables and of factors of different degrees of generality. Guilford suggests, as a remedial strategy, that we ought ideally to start at the lower levels and work our way upward. Since no one has yet followed such a plan consistently, it is difficult to say how practical it would prove. Perhaps it would prove more feasible for some areas of behavior than for others. In any case, we are still faced with the problem of determining more objectively the measurement level of whatever variables we consider and obviating the sort of guesswork in which the writer has just indulged.

Assuming that our researches will ultimately cover the entire continuum of referent generality, perhaps what we need in any single investigation is not so much low generality sampling of variables as uniform generality and uniform density sampling. The aim of such sampling, of course, would be to obtain factors of relatively uniform generality. The desired generality of either variables or factors may best be determined by our immediate needs for scientific description. It is unlikely that a sound procedure for satisfactory sampling can be prescribed at this time, and the problem is likely to require a considerable amount of exploratory work with all sorts of measurement variables. A possible solution might lie in selecting variables out of a much larger potential set on the basis of their patterns of correlation. For example, if an initial large sample of variables were designed to cover very intensively a certain band within the hierarchical system for a given realm of observation, we might select variables for factorization on the basis of the diffusiveness of their correlations with other variables. The hierarchical model assumes implicitly that the higher the level of generality, the more the variable will tend to display moderate correlation with many other variables and very high correlation with few other variables. The major obstacle in the way of applying such a procedure is the difficulty of satisfying other requisite assumptions regarding the nature of initial sampling and the uniformity of mode of organization within the total system.

Perhaps further exploration will suggest entirely different modes of attack on the whole problem. Cattell has long been an advocate of comprehensive sampling of variables within any realm one chooses to study, and the advent of computers has helped to make this aim appear reasonably attainable. In terms of our hierarchical model, we can now handle samples of great horizontal scope. Perhaps in this age of computers we can also find a basis for refining our selection in terms of density and generality. A methodological and empirical foundation will then have been laid for constructing an integrated picture of human behavior and experience that will encompass both the most general tendencies and the most specific actions.

It will probably occur to some readers that Guttman's facet theory (cf. Guttman, 1954–55, 1958) provides the basis for a different approach to the problem of uniform generality sampling—particularly since Humphreys (1962) has recently advocated the facet model as a "more general way of thinking about psychological tests" than the hierarchical model. Humphreys argues that the hierarchical model breaks down in the case of the multitrait-multimethod matrix of Campbell and Fiske. The hierarchical model "breaks down" in a multifacet system, of course, only in the sense that the subordination pattern within the hierarchy becomes complex. Furthermore, one may question whether the facet model is really an alternative to the hierarchical model, since it seems to serve a different purpose and to rest on a different logical basis. Facet theory is based on a logical analysis of test structure, whereas the hierarchical model (as conceived by Burt) rests on a logical analysis of trait distributions (more precisely, in its implementation, the distribution of traits over a sample or population of behaviors).

Facet theory affords a basis for its own species of hierarchy, one in which levels correspond to orders of interaction between the facets of a design. Such a hierarchy is valuable in its own right, but the kind of generality that is thus represented is logically distinct from both contextual generality and referent generality. Even when test sampling is deliberately "heterogenized" with respect to facet combinations in a given research, there is nothing to guarantee a close correspondence between facet interaction level and the other kinds of generality. Guilford's (1956, 1959b) recent classification of intellectual factors may be construed as representing an alignment at one interaction level of factors that are hopefully of relatively uniform referent generality. We may well find in future research, however, that the structure that proves most empirically meaningful at higher levels of referent generality in the realm of intellectual performance bears little resemblance to anything suggested by the facet model.

Before dismissing facet theory as irrelevant to our purposes, we should recognize the advantages it would offer if referent generality could somehow be linked to the facet model. If and where such a correspondence can be established, the facet model would provide a basis for identifying referent generality independently or semi-independently of contextual generality, permitting allowances for the peculiarities of variance distributions in specific sets of data. Such an eventuality may depend on the development of appropriate techniques for identifying facets and elements within them. Basically, the problem of uniform generality sampling remains an empirical problem which can best be conceptualized in terms of a modified hierarchical model. Humphreys, however, has performed a valuable service in suggesting the relevance of facet theory to this model.

REFERENCES

ALLPORT, G. W. What is a trait of personality? *J. abnorm. soc. Psychol.*, 1931, **25**, 368–372.

ALLPORT, G. W. *Personality, a psychological interpretation.* New York: Holt, 1937.

ANASTASI, ANNE. Faculties *versus* factors: A reply to Professor Thurstone. *Psychol. Bull.*, 1938, **35**, 391–395.

BARTLETT, M. S. The statistical conception of mental factors. *Brit. J. Psychol., Gen. Section,* 1937, **28**, 97–104.

BURT, C. *The factors of the mind.* New York: Macmillan, 1941.

BURT, C. The structure of the mind: A review of the results of factor analysis. *Brit. J. educ. Psychol.*, 1949, **19**, 100–111, 176–199.

CATTELL, R. B. *Personality and motivation structure and measurement.* Yonkers-on-Hudson: World Book, 1957.

CATTELL, R. B. Theory of situational, instrument, second-order, and refraction factors in personality structure research. *Psychol. Bull.*, 1961, **58**, 160–174.

CATTELL, R. B., & SULLIVAN, W. The scientific nature of factors: A demonstration by cups of coffee. *Behav. Sci.*, 1962, **7**, 184–193.

COAN, R. W. A comparison of oblique and orthogonal factor solutions. *J. exp. Educ.*, 1959, **27**, 151–166.

COAN, R. W. Theoretical concepts in psychology. *Brit. J. statist. Psychol.*, 1963, **16**, in press.

EYSENCK, H. J. *Dimensions of personality.* London: Routledge & Kegan Paul, 1947.

EYSENCK, H. J. The organization of personality. *J. Pers.*, 1951, **20**, 101–117.

GUILFORD, J. P. Human abilities. *Psychol. Rev.*, 1940, **47**, 367–394.

GUILFORD, J. P. The structure of intellect. *Psychol. Bull.*, 1956, **53**, 267–293.

GUILFORD, J. P. *Personality.* New York: McGraw-Hill, 1959. (a)

GUILFORD, J. P. Three faces of intellect. *Amer. Psychologist,* 1959, **14**, 469–479. (b)

GUILFORD, J. P. Factorial angles to psychology. *Psychol. Rev.*, 1961, **68**, 1–20.

GUTTMAN, L. An outline of some new methodology for social research. *Publ. Opin. Quart.*, 1954–55, **18**, 395–404.

GUTTMAN, L. What lies ahead for factor analysis? *Educ. psychol. Measmt.*, 1958, **18**, 497–515.

HUMPHREYS, L. G. The organization of human abilities. *Amer. Psychologist,* 1962, **17**, 475–483.

KELLEY, T. L. *Essential traits of mental life.* Cambridge: Harvard Univer. Press, 1935.

MOURSEY, E. M. The hierarchical organization of cognitive levels. *Brit. J. Psychol., Statist. Section,* 1952, **5**, 151–180.

SCHMID, J., & LEIMAN, J. M. The development of hierarchical factor solutions. *Psychometrika,* 1957, **22**, 53–61.

SPEARMAN, C. *The abilities of man, their nature and measurement.* New York: Macmillan, 1927.

THOMSON, G. H. On complete families of correlation coefficients, and their tendency to zero tetrad differences: Including a statement of the sampling theory of abilities. *Brit. J. Psychol.*, 1935, **26**, 63–92.

THOMSON, G. H. *The factorial analysis of human ability.* London: Univer. London Press, 1951.

THURSTONE, L. L. *The vectors of mind, multiple-factor analysis for the isolation of primary traits.* Chicago: Univer. Chicago Press, 1935.

THURSTONE, L. L. Shifty and mathematical components, a critique of Anastasi's monograph on the influence of specific experience upon mental organization. *Psychol. Bull.*, 1938, **35**, 223–236.

THURSTONE, L. L. *Multiple-factor analysis.* Chicago: Univer. Chicago Press, 1947.

THURSTONE, L. L. Psychological implications of factor analysis. *Amer. Psychologist,* 1948, **3**, 402–408.

TRYON, R. C. A theory of *psychological* components—an alternative to "mathematical factors." *Psychol. Rev.*, 1935, **42**, 425–454.

TRYON, R. C. General dimensions of individual differences: Cluster analysis vs. multiple factor analysis. *Educ. psychol. Measmt.*, 1958, **18**, 477–495.

VERNON, P. E. *The structure of human abilities.* London: Methuen, 1950.

(Received February 4, 1963)

Psychological Review
1964, Vol. 71, No. 2, 141–147

A METHOD OF ANALYSIS AND CLASSIFICATION OF REPETITIVE RESPONSE SYSTEMS[1]

ROBERT B. LOCKARD

University of Washington

Many animal response systems are characterized by a stable rate component. For example, daily caloric intake is stable in a stable environment. Many such response systems supply reservoirs (e.g., drinking) with a characteristic drain rate. Thus a "theoretically perfect" reservoir system would accumulate a deficit when denied inflow (e.g., denied drinking behavior) but would exactly compensate for the accumulated deficit when inflow was again made possible. This basic model is extended to all repetitive response systems, including those without known reservoirs, and those which do not compensate for deprivation. Thus a quantitative taxonomy of diverse behaviors becomes possible.

It is not the purpose of this paper to propose a method for the analysis of learning curves. These, broadly speaking, pertain to response systems in transitional states as a consequence of a change in the functional properties of the environment. The class of phenomena to which this paper applies is stabilized response systems operating in a constant environment. Some of these systems are so regular with respect to response frequency per unit of time that rates are easily computed and are useful as dependent variables or as indicators of other events. For example, we speak of heart rate and respiratory rate. Certain other response systems, such as activity-wheel running rate, vocalization rate, and the rate of other striped muscle activities, are not in any obvious way a part of conservative homeostatic systems. Yet many such systems display a temporal regularity over the appropriate time base. It is the purpose of this paper to show that many repetitive response systems can be analyzed by the same set of operations. Two quantitative system parameters are derived from the way and extent that a given system responds to deliberately introduced perturbations. Since numerous response systems can be measured in the same way, a quantitative taxonomy becomes possible. This general approach might be called "behavioral statics"—the analysis of behavior before or after a learning situation, as contrasted with the traditional concern of altering behavioral rates through the introduction of contingencies.

One class of response systems ultimately involves a reservoir of some kind. Breathing, eating, and drinking might be said to fill reservoirs. Urination and defecation drain upon reservoirs. The behavior associated with each of these systems is identifiable and repetitive; the animal breathes, eats, drinks, and defecates repeatedly. These different responses are also characterized by rates. When cumulative responses are plotted as a function of time, the rate of the ongoing behavioral system appears as the slope of the function. If the animal is deprived of the opportunity to respond by plugging the nostrils, withdrawing food or water, or closing the urethra, the response rate becomes zero. When

---

[1] The author wishes to express an intellectual debt to William S. Verplanck, who originally conceived the fundamental ideas this paper merely developed.

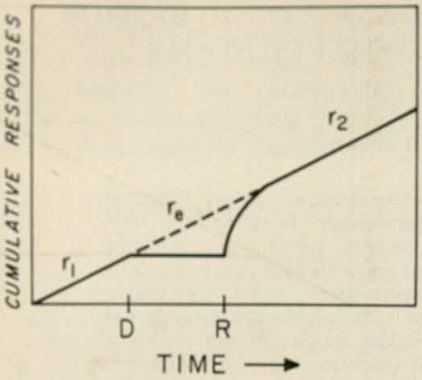


FIG. 1. Example of a "theoretically perfect" response system. (The characteristic response rate, r_1 was forced to zero at time *D*. At Time *R* the opportunity to respond was restored. The system compensated for the deficit accumulated during deprivation, then returned to its characteristic rate.)

the opportunity to respond is restored, what might be called "replacement behavior" takes place; there is an initial high rate of response which continuously decreases until, at some later time, the response rate stabilizes. This typical course of events is illustrated in Figure 1. The predeprivation rate, r_1, is forced to zero by experimental procedures at the point in Time *D* (for deprived). When the opportunity to respond is restored (at *R*), replacement behavior occurs. The system reaches rate stability r_2, which is depicted as being equal to r_1. At the time of restoration of the opportunity to respond, reservoir systems would have a theoretical deficit (need) equal to the difference between r_e, an extrapolated (or control group) rate, and the response function at Time *R*. The deficit results from the organism's drain upon material in the reservoir, and replacement behavior serves to restore the reservoir to its normal psychological parameters. Thus Figure 1 depicts a reservoir system as showing complete, or 100% replacement.

Certain other response systems, though repetitive and stable in rate, serve no known reservoir function. Examples of these behavioral systems would be the manipulation behavior of monkeys, running wheel activity of rodents, and saccharine ingestion. Figure 2 illustrates the response of the hypothetical nonreservoir system to the same deprivation procedures employed in the first example. At Time *D*, the animal was deprived of the opportunity to respond, thus forcing the response rate to zero. After an interval of deprivation, the opportunity to respond was restored at Time *R*. In this example, replacement behavior occurred but was not complete. The theoretical deficit at Time *R* was replaced by 20%. At about Time *T*, r_2 stabilized and became equal to the predeprivation rate r_1, as their equal slopes indicate. The deficit, however, was never further diminished. This illustration suggests that behavioral systems are not properly dichotomized into replacement and nonreplacement (or drive and nondrive) systems, but that replacement is a quantifiable continuous variable.

Those behavioral systems without known reservoirs and which allow sur-

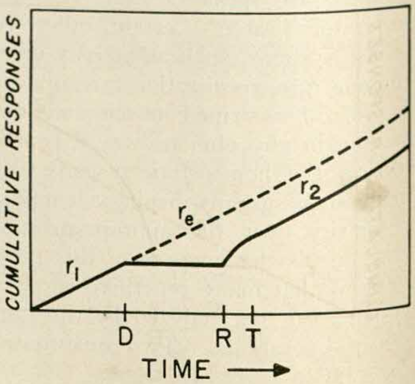


FIG. 2. Example of a repetitive response system which, following deprivation, partially compensates for the theoretical accumulated deficit.

vival following infinite deprivation intervals are those which have the most controversial drive status. One typical experimental procedure employed to test for drivelike effects of an unknown system allows a deprived experimental group and an undeprived control group to respond again at Time *R* of Figure 2. At Time *T*, or after some arbitrary interval, the responses of the two groups are totaled, and it is often found that the experimental group responded more than the undeprived group. This classical "test for drive" can discover but one of two nominal classifications for response systems—drive and nondrive. As Figure 2 shows, a system showing any appreciable amount of partial replacement might be clasified as a "drive" system without discriminating between 100% replacement systems and low replacement systems. The status of presently controversial behavioral systems might be clarified by procedures which measure the amount of replacement behavior on a continuous scale with more than two categories.

Certain replacement behaviors are more than complete. Hoarding (e.g., Morgan, 1947) following deprivation serves as an example, and certain complex human responses are sometimes said to be overcompensatory following deprivation. Overcomplete replacement behavior is illustrated in Figure 3, which shows 160% replacement following deprivation. In this hypothetical example, r_1 was the daily rate of food-transport to the home cage prior to deprivation. After the opportunity to respond was restored (Time *R*), more food was transported to the home cage than would have been (r_e) had there been no deprivation.

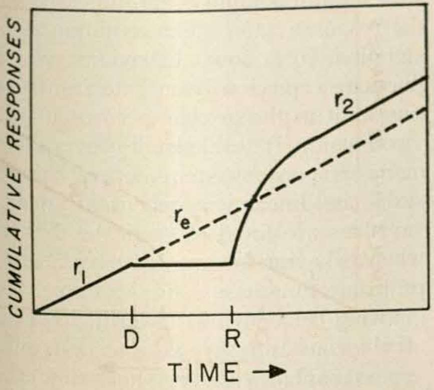


Fig. 3. Example of a repetitive response system which, following deprivation, overcompensates for the theoretical accumulated deficit.

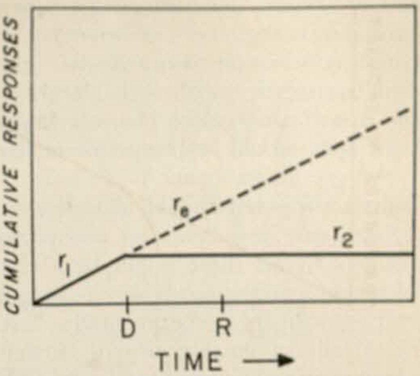


Fig. 4. Example of a bimodal reversible response system. (When the opportunity to respond was restored at Time *R*, the system did not return to its "addicted" rate r_1 but assumed the preaddiction rate r_2. The system's two operating modes are r_1 and r_2. The system is reversible because it can be unaddicted.)

The several hypothetical examples already discussed were assumed to return to the predeprivation rate after the system was experimentally disturbed. Deprivation, however, might alter the rate parameter of certain repetitive response systems such as addicted behavior. Figure 4 shows a hypothetical addiction system (e.g., smoking) operating at rate r_1. Following deprivation, r_2 is zero. It can be safely assumed that some previous rate was zero, but, following loading procedures, became stabilized at r_1. Since deprivation restored the functional properties of the behavioral system to a former value (not smoking), the system could be termed reversible.

A general taxonomic device such as is here proposed should allow for the general case, and hence for unexpected results. Thus there might be a class of response systems whose rate parameter r_2 can be experimentally "set" at various values intermediate between abstinence and some upper limit by the operations of deprivation and loading. A change in the rate parameter of a response system can be expressed as

$$\Delta r = \frac{100\left[\arctan \frac{r_2}{r_1} - 45\right]}{45}$$

in which Δr is the change in rate, r_1 is the predeprivation or preloading rate, and r_2 the postdeprivation or postloading rate. The effect of the several constants is to create a scale ranging from 100 to −100, with zero representing no change in rate, or $r_2 = r_1$.

The other system parameter, percentage of replacement C (for compensation), can be clarified by reference to Figure 2. The theoretical deficit at Time R is the vertical difference between r_e and the response function. At some Time T, r_2 has stabilized and the deficit has been replaced by 20%. C is thus defined as

$$C = \frac{d_R - d_T}{d_R},$$

with d_T equal to the theoretical deficit at time T and d_R the deficit at time R. C can range from zero to ∞, with $C = 100$ representing complete replacement of the theoretical deficit, as depicted in Figure 1.

The two parameters C and Δr suffice to describe the response to experimentally introduced perturbations of behavioral systems as diverse as eating, drinking, manipulation, locomotion, hoarding, self-injected drugs, etc. These parameters are quantitatve, their relevant operations are specified, and they are independent of the type of response, response units, and time units of a given experiment. The four hypothetical examples can be adequately described by their C, Δr values. The ideal reservoir system of Figure 1 is a 100, 0 system; the partial replacement system of Figure 2 is a 20, 0 system; and the next two figures show 160, 0 and 0, −100, respectively.

A graphic summary of response systems within a parameter space is possible. With Δr on the abscissa and C on the ordinate the two parameters of a given system locate it as a point in the two-dimensional space. For example, reservoir systems would probably cluster together in the central region of 100, 0. Other clusters might occur as natural groupings in other locations, and it would be possible that systems within a cluster showed similar motivational effects.

The illustrations discussed have assumed both stability and linearity of the response rate. The assumption of stability rules out behaviors which fluctuate capriciously in rate from one time unit to the next in a constant environment. It is likely, however, that many response systems operate in a stable and linear way when extraneous variables are fixed at static levels and when behavior is sampled over an appropriate time base. For example, the gnawing behavior of rats (which wears off the constantly growing teeth) could appear haphazard over short time periods but reveal order over long time samples.

The assumption of linearity is a

first approximation. If a system is stable but nonlinear, extrapolations for the special case could be made. In these cases, a control group is especially suited for determining r_e. It might be noted, however, that any constantly accelerating rate function means that the behavior in question would soon either vanish or, for increasing rates, come to occupy all the organism's time. Thus it is reasonable to expect that most systems would demonstrate but small departures from linearity, possibly with an order of magnitude small with respect to experimental error.

In the examples discussed no specific deprivation interval was assumed. There are at least two major considerations in this regard. The first is that the interval should be appropriate to the normal timebase of the system. For example, animals cannot be air-deprived for as long a time as they may be water-deprived. This consideration is probably encompassed within the second, which stems from the evidence that replacement behavior is not always a linear function of time. For example, Stellar and Hill (1952) measured water replacement behavior following various deprivation intervals. Their data indicate approximate linearity of replacement up to about 48 hours of deprivation. It seems quite probable that many response systems, as deprivation continues, react by altering internal system parameters. To the extent that these parameters are altered, the system is no longer the "same" system as entered deprivation. Inanition, stomach shrinkage, greater or lesser activity, or sensory adaptation could change the system parameters underlying r_1. Therefore, the behavior of control groups or extrapolations of r_1 would not be legitimate data for inferring accumulated deficits. Until more is known about the nonlinear properties of response systems during exaggerated deprivation, the exprimental operations suggested in this paper would probably be best confined to that portion of the deprivation interval during which replacement behavior is a linear function of time.

A POSSIBLE TAXONOMY OF REPETITIVE RESPONSE SYSTEMS

The two parameters C (percentage of replacement behavior) and Δr (change of response rate) were suggested as adequate for locating a given response system in two-dimensional space. Such a space is shown in Figure 5, and located in it are the figure numbers of the four hypothetical systems used as illustrations. The location of numerous other systems could have been guessed at by the writer, but the reader may do this for himself without publishing his errors. As an example of locating an unknown system, one might imagine that, following deprivation of bathing, a human would show some slight replacement behavior followed by the same bathing rate exhibited prior to deprivation. This response system would then be located in Figure 5 in the region of perhaps 10, 0. The system also bears extension beyond the animal kingdom; for example, the number of cubic feet of water flowing past a point in a stream bed could be defined as "responses." A temporary dam is built to create a deprivation interval, then opened to allow "the opportunity to respond." Simple streams, without alternate overflow paths, would doubtlessly prove to be 100, 0 systems.

In addition to locating response systems in C, Δr space, various nominal classifications can be anticipated. Rate-stable, or *unimodal,* systems

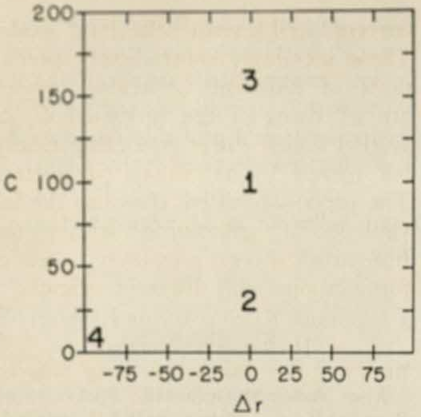


FIG. 5. Example of locating response systems in the two-dimensional space described in the text. (The located numbers refer to figures illustrating the system.)

would be those capable of assuming but one rate under standardized conditions. Breathing serves as an example. *Bimodal* systems would be those capable of operating in two modes, such as a system subject to addiction. *Polymodal* systems would be those capable of being "set" at various rates by experimental procedures. All polymodal systems would occupy more than one location in C, Δr space, and different behaviors would correspond to the system locations. For example, an opiate might be expected to be a reinforcer for an addicted animal but not for an unaddicted one. *Reversible* polymodal systems would be those which could be returned to Mode 1 after being set in Mode 2, as in breaking a habit by deprivation. *Symmetrical* systems would be those yielding equal C, Δr values following both deprivation and loading. As an example of possible taxonomy, the water-ingestion system might be a unimodal irreversible symmetrical 100, 0 response system. In this particular case, "irreversible" is redundant, for bimodality is a necessary condition for reversibility.

Some response systems are obviously *surrogate*. For example, a rat deprived of responding to drinking tube *A* will show little or no replacement behavior if drinking tube *B* remains during the deprivation interval. Response systems *A* and *B* are thus totally surrogate when tubes *A* and *B* both deliver water. System *B* could be made partially surrogate for *A* by adulteration, for example with quinine sulphate. Whenever a response system shows partial replacement, surrogate systems should be sought. For example, home cage activity may be partially surrogate for activity-wheel locomotion, and an animal deprived of the latter might show an increase in the former which, unless instrumented, would go undetected. At present, practically nothing is known of the surrogate properties of various response systems although much behavioral methodology is based upon the ability of organisms to substitute one response system for another. For example, a rat deprived of the opportunity to respond for food via System *A* must substitute System *B* (e.g., bar pressing) to maintain r_e, here defined as caloric intake. But since psychological interest has focused on systems in transition, as opposed to the properties of steady-state systems, response substitution phenomena have not usually been viewed as such. Had they been, certain past findings might have been viewed more thoroughly in this framework. For example, visual exploration responses of monkeys (Butler, 1953) could have stimulated investigation of possible home cage systems for which door opening substituted, to what degree they were surrogate, and so forth. A taxonomy of response systems, based upon analysis of their modal, symmetrical, and surrogate properties, might do much to clarify the problem of one *D* versus many drives (e.g., Brown, 1961) and would seem relevant to issues regarding displacement behavior.

The method of analysis proposed herein is not a theory of behavior. It suggests a set of operations for the quantification and subsequent classification of response systems. It is suggested, however, that systems found functionally similar will also be found similar in new behavioral situations, especially those to which concepts of motivation apply. At present, there are two seemingly antithetical behavioral models clearly applicable to postdeprivation behavior: the homeostatic model discussed by Young (1951) and the adaptation-level model of Helson (1959). Following deprivation of an unknown Response System *X*, the homeostatic model predicts high replacement behavior and no change in response rate. The adaptation-level model predicts low replacement behavior and $r_2 < r_1$. Another likely outcome, low replacement behavior and no change in rate (e.g., Figure 2), is not a prediction of any current, well-known behavioral model. These seemingly contradictory predictions of different behavioral models are all cases of the proposed C, Δr system which differ not categorically, but merely in continuous parameters. The proposed system thus has the advantage of subsuming within one framework several seemingly unrelated formulations with different origins.

REFERENCES

Brown, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

Butler, R. A. Discrimination learning by rhesus monkeys to visual-exploration motivation. *J. comp. physiol. Psychol.*, 1953, **46**, 95–98.

Helson, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959.

Morgan, C. T. The hoarding instinct. *Psychol. Rev.*, 1947, **54**, 335–341.

Stellar, E., & Hill, J. H. The rat's rate of drinking as a function of water deprivation. *J. comp. physiol. Psychol.*, 1952, **45**, 96–102.

Young, P. T. Motivation of animal behavior. In C. P. Stone (Ed.), *Comparative psychology.* Englewood Cliffs, N. J.: Prentice-Hall, 1951.

Psychological Review
1964, Vol. 71, No. 2, 148-159

A COUPLED SYSTEM FOR CODING AND LEARNING IN SHAPE DISCRIMINATION

P. C. DODWELL

Queen's University, Kingston, Ontario

A system for coding visual information about shape is described. It is shown how the output from this coding system can be coupled to a classifier, or recognizer, in such a way that shape equivalence is preserved. A major weakness of previous coding-type theories of shape recognition is remedied, in that a memory system is built into the recognizer, which becomes self-organizing in terms of stimulus regularities in the environment. This is shown to be a logical basis for perceptual learning and perceptual constancy. Recent confirmatory neurophysiological findings from direct recordings in the striate cortex of cat are briefly considered, and the possibility of coupling the recognition system to a response system is outlined.

INTRODUCTION

It has been argued elsewhere (Dodwell, 1961c) that the orthodox probabilistic approach to constructing theories of discrimination learning is weak, because of the ambiguity with which the term "stimulus" is used in such theories. An ambiguous term in an abstract theory may have functions similar to those of a *deus ex machina,* and should be regarded with suspicion. It was also argued that a more precise definition of the term was needed, and that one way of defining stimulus would be in terms of a system for coding or classifying sensory input. Indeed, for some types of visual discrimination, such as shape recognition, there is now considerable behavioral evidence to indicate the general principles on which such classifying systems may work in fairly simple visual systems (e.g., Deutsch, 1960; Dodwell, 1957, 1960, 1961a, 1961b; Sutherland, 1957, 1959, 1960, 1963). This evidence has recently been supplemented by neurophysiological findings obtained by direct recording from individual neurons in the visual cortex of cat (Hubel & Wiesel, 1959, 1962).

The concept of shape coding, as originally suggested by Deutsch (1955), can be held to be unsatisfactory because it does not allow for the operation of perceptual learning. This criticism can be leveled at a shape coding theory as a theory of shape recognition, that is, as a *complete* theory about how shapes are recognized, but there is no reason why the output from a coding system should not be coupled to a learning system or "response probability generator" to yield a more complete and logically satisfactory theory. In fact it is the purpose of this paper to show that it can be done, and to show that the coupled system is not open to the sorts of criticism which can be made of either component (i.e., the shape coding system and the probabilistic discrimination learning model) on its own.

The system to be described consists of an elaborated version of a shape analyzing system proposed earlier to explain shape discrimination in rats (Dodwell, 1957), connected to a signal classification system suggested by Uttley (1954). This dual system does not directly predict response probabilities in the way stochastic models do, but a straightforward extension of the double system can do so. This combined model gives a good fit to most of the behavioral data on shape discrimination in the rat and octopus, and also incorporates many of the features discovered in the direct neurophysiological exploration of the striate cortex of cat.

THE SHAPE CODING SYSTEM

The system envisaged consists of a two-dimensional array of units arranged in parallel chains, so that each unit has two neighbors, except for the end units of each chain, each one of which has only one neighbor. The chains are functionally independent except in two important respects: one set of end units (Set A) is so arranged that if one of them fires, they all fire simultaneously, and each one of the other set of end units (Set B) is connected to a single transmission line, called the final common cable (fcc). Each chain has the same number of units (this is not a necessary condition of the working of the system, but its introduction simplifies the exposition). Each unit is connected to a receptor, or group of receptors, on the retina, and when the receptor (or group of receptors) is stimulated this activates the unit to which it is connected. Such a unit is "active"; when a unit is not so activated it is "passive." An active unit does not affect any of its neighbors by transmitting "activation" to them. It is assumed that only excitation from regions of sharp change in stimulation (contours) activates the array, and that adjacent receptors (or groups of receptors) activate adjacent units. Thus, when the image of a sphere is focused on the retina, a set of units corresponding to a circular contour is activated on the array. "Corresponding" is used advisedly since, although a one-to-one correspondence between receptors (or groups of them) and units is postulated, this does not imply any simple geometrical similarity in the mapping of the points in the two regions. However, an important assumption is made about horizontal straight lines: it is assumed that (with the retina in normal orientation) stimulation by a horizontal line will activate units all on the same chain.

A second process in the units, called *excitation*, originates in the firing of the End-set A, and results in a "sweep" across the array, since each unit can pass excitation from itself to its neighbor in the direction of the End-set B. All units of this set discharge into the fcc. This excitation is not to be confused with activation of units by receptors, or groups of receptors, to which they are connected. The transmission of excitation from unit to unit involves a small delay, the length of which is determined by whether the transmitting unit is active or passive. Only the two states are envisaged, and the delay for each is constant, say Ta and Tp. For convenience let $Ta > Tp$.

Each unit can pass excitation to one neighbor only, and excitation always travels from A towards B. The time for excitation to traverse a chain of $n + 1$ units, all of which are passive, is $n\ (Tp)$. If the units of the array are passive, the traverse times will be equal, and there will be a single large discharge into the fcc, since all the A units fire together and excitation reaches all the B units simultaneously.

When some units are active, some passive, the traverse time for any particular chain will depend only on the number of active and passive units on it, and can be written in the form $n\ (Ta) + (n - r)(Tp)$. The added delay involved when r units are active on a chain is obtained by subtracting

$n\ (Tp)$ from this expression:

$$r(Ta) + (n - r)(Tp) - n\ (Tp) = r(Ta - Tp)$$

Let the additional time for transmission through an active unit be ΔT. Thus $\Delta T = Ta - Tp$, and the added delay on a chain with r active units is $r\Delta T$. This may be called the "delay function" for a single chain. The delay function for an array of N chains will be made up of a set of terms of this form. Effectively, the system transforms spatial information into a function of a single variable, t. Thus, the characteristics of the delay function are independent of which particular chains, or which units on those chains, are activated. That is, the signal in the fcc is independent of the signaled shape's position (but not its orientation) on the array.

Sets may be formed for chains having equal values of r. One of these will be the set for which $r = 0$, which may be written $[q_o,\ nTp]$, where q_o represents the number of chains in the set, and hence the amplitude of the initial volley in the fcc. The total signal, for one "sweep" will be the *union*, or logical sum, of these sets. The set for any one value $r = r_k$ contains two components, q_k, and $r_k\Delta T$. The former is a measure of the amplitude of pulse generated in the fcc, the latter a measure of the time at which the pulse is generated, measured from the initial pulse (when $r = 0$) as base line. The delay function for the array is sufficient for deducing the important characteristics of a signal if, as would often be the case, the number of empty chains and the size of the initial volley are unimportant. In general it is necessary that there should be either at least one large set with equal r, or at least one chain with a value of r that is atypical, for a distinctive signal to be generated (either q_k or ${}_kr$ or both, identify a set).

It has already been stated that the array is assumed to be so arranged that chains represent, functionally, one of the principal axes of the visual field; stimulation of receptors by a horizontal line will activate units all on one chain. Which chain this is will be immaterial as far as the signal in the fcc is concerned; the delay function is simply $[1,\ r\Delta T]$, and the distinctiveness of the signal is due to this one atypical value of r (for all other chains, $r = 0$). Two horizontal lines would be distinguished from one by a doubling of the amplitude of the delayed discharge, and so on.

In general, the system will yield distinctive signals for different shapes, and is sensitive to changes in orientation (rotation in the frontal plane). Simple geometric considerations show that this sensitivity is greatest for small rotations of a horizontal straight line (provided the density of units on the array is uniform) whether or not this is part of a shape, such as the base of a square or triangle. The same considerations show that it is least sensitive to rotations near the vertical. However, as it stands the system will not yield distinctive signals for some pairs of shapes which one would expect to be readily discriminable, such as a diamond and a pair of vertical lines. This difficulty can be overcome by postulating two systems, functionally perpendicular to each other, so that in the second a vertical line in the visual field activates units all on one chain. Such a dual system will explain fairly satisfactorily most of the data on shape discrimination in the rat and in the octopus. For instance, it will explain the fact that both species can readily discriminate horizontal from vertical straight lines, but have difficulty with mutually perpendicular lines set at 45° to the horizontal; the fact that mirror images tend to be confused, that transfer tends to be easy (for the rat) when the transfer shape shares a horizontal contour with the training shape and that rotation of shapes in the sagittal plane tends to upset discrimination. Detailed evaluations of the empirical evidence for this, and for other shape coding systems, have been published elsewhere (Dodwell, 1957, 1961b) and will not be repeated here. While it is true that some facts about shape recognition are not compatible with the theory, it should be remembered that other means of discrimination, such as by brightness distribution, may enter as contaminating factors (see Dodwell, 1961b). The important feature to be brought out at the moment is that the primary information processing postulated in the coding system is the first step in the specification of a logically satisfactory theory of shape recognition. The features of the system that are important in this respect are, first of all, the fact that in it spatial information is transformed into a temporal analogue, and secondly, that this analogue consists of sequences of pulses of different amplitudes and latencies which all occur in a single transmission line. It is not claimed that this is the only suitable way of coding information about shape for further processing, but it has two major advantages.

The first advantage is that the information about shape is coded independently of the position on the array (and hence on the retina) at which the shape was present. This means that the system overcomes the problem of identification of a shape in different locations—which has always been a major difficulty for neural theories of shape discrimination—in a simple fashion. The second advantage is that it makes possible an attempt to solve the problem of stimulus equivalence in general (recognition of shape independently of position is, of course, also part of the problem of stimulus equivalence, but is a rather special case). Explicit separation of the operations of "coding" and "recognizing" allows one to reformulate the problem of stimulus equivalence into the question: which outputs from the coding system are classified in the same way by the "recognizer"? To answer this question one must consider how the recognizer operates on the outputs from the coding system.

THE RECOGNIZING SYSTEM

How can "recognition" be defined in a formal system of the type here envisaged, which is to be equated, in some sense, with neural processes in the brain? One may define recognition as the registration of some constant pattern of activity, or activity at some constant locus, for a given input pattern or class of input patterns. Such a system will now be described.

The recognizer accepts and classifies the output from the coder, which in the present system consists of sequences of pulses of varying frequency and amplitude, all occurring in a single transmission line (or in two lines for the double system; the single system will be described, since the problem is in principle identical, but more complicated, for the double system). A system has been outlined by Uttley (1954) which, with slight modifications, will classify inputs of this type. As it stands, Uttley's system classifies temporal patterns of discrete pulses of equal amplitude, and is thus a digital system. The outputs of the coding system described above are analogues of the geometrical properties of shapes; however, Uttley (1954) indicates a simple analogue-

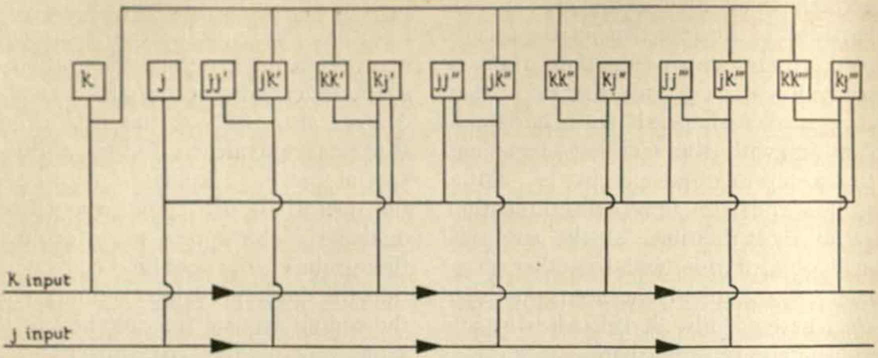


FIG. 1. The recognizer for outputs from a shape coding system. (For explanation, see text.)

digital conversion system, of a type which might be used in a computer, which could occur in a nervous system (p. 486). There is evidence that analogue-digital conversion does occur in certain neural structures (e.g., Hilali & Whitfield, 1953), so it is not implausible to suggest that some such conversion occurs between the outputs of the coder and the recognizer. The particular form of analogue-digital conversion suggested requires that the recognizer have a number of inputs, and is as follows: each volley in the coder output gives rise to a "unit" impulse in one of the inputs to the recognizer. If l different amplitudes are to be discriminated, there must be l inputs in the recognizer: each input fires for one and only one value of amplitude q_k, or for only a small range of such values. For the simplest case, where only two different amplitudes are recognized, two inputs to the recognizer are required, inputs j and k, say. These two inputs are connected to counters which register either (*a*) the initial arrival of unit impulses in either input, or (*b*) the temporal patterns of impulses on either or on both inputs. This is achieved by feeding each input into a series of delay units, and then to the counters, as shown in Figure 1. (The recognizer can be imagined as either a nerve network or as an electrical system; however, its properties are specified independently of any particular embodiment.) Except for the first two counters (j and k), a counter will only register when two impulses arrive at it simultaneously or within an arbitrary small interval of time. To see how this occurs, consider the counter labeled jj'' in Figure 1. It will register if, and only if, an impulse in the j input is followed by a second impulse in the same input after time τ, where τ is the delay induced by the first delay line on input j. Similarly, the adjacent count jk' registers only when an impulse on j is followed by an impulse on k after time τ. (The primes refer to the number of delay units separating a pair of impulses.) Assuming $\tau = h\Delta T$, where h is an integer, the time intervals required to activate different counters will be simply related to the output from the coder. It can be seen that the recognizer of Figure 1 will discriminate all possible patterns for the two inputs j and k, up to time delays of 3τ. For patterns involving more than two impulses, more than one counter (excluding the initial j and k counters) will register.

A simple higher order type of recognition can easily be added, whereby a counter $jj'kk''$, for example, will only register when both jj' and jk'' are registering. jj' and jk'' fire simultaneously if, and only if, an impulse in k is followed by one in j after time 2τ, followed by a second impulse in j after a further time τ. In this case the counter jk''' would also fire, of course, so that at a third level of classification a counter $jj'jk''jk'''$ would register. This hierarchical system of classification can be continued to any desired level of complexity, according to the number of inputs, and the number and complexity of patterns to be classified. The only restriction required is that $u\tau < n\Delta T$ where U is the number of delay units in the recognizer, and n, as before, is the number of units per chain in the array.

"Equivalent stimuli" can now be defined as inputs from the receptors which give rise to the activation of the same set of counters in the recognizer. It may be thought that little is gained by introducing a recognizer of this form—that in fact the original information about shape has been lost almost beyond recovery, and that no compensating advantage has been secured. However, it should be noted, in the first place, that original signal information is contained unambiguously in the system of registers that are activated: if the system were reversed, so that the counters became signal generators, the original input signal would be perfectly reproduced for a given constellation of registers, but now as an output from the system (Uttley, 1954). Secondly, a distinct advantage has been secured in the sense that a structure is set up by which a constant local pattern of activity is a necessary consequence of a spatial pattern, at the receptors, which is *not* location specific. As Ayer (1957) points out, the occurrence of a necessary consequence is in all respects equivalent to a sufficient condition for the antecedent's occurrence; i.e., the occurrence of a given pattern at the recognizer is sufficient to ensure that a particular pattern *did* activate the receptors. The only exception here is the fact that certain different patterns may give rise to identical or nearly identical patterns from the coder, in which case perceptual confusion arises. The question of whether or not "perceptual confusion" is to be equated with "perceptual equivalence" is taken up in the next section of the paper. Before taking up this question, it may be noted that an information-reducing system such as the coder, interposed between the receptors and the recognizer, serves to remove an objection that has been raised against the Uttley-type recognizer. This is the objection that, to be effective, such a classificatory system would need an impossibly large number of units to deal with all combinations of receptor activation patterns (Uhr, 1963).

PERCEPTUAL LEARNING

The coupled system so far described has no "memory"; its performance is entirely independent of its own previous states. As such, therefore, it is deficient as a model for perception. To show how the system can be modified to incorporate a type of perceptual learning, the notion of a conditional probability computer can be introduced. The idea is, again, due to Uttley (1956, 1958, 1959). Consider the hierarchical classification system of Figure 2. The lines a, b, and c represent inputs, the open circles represent, as in Figure 1, coincidence counters. Simultaneous inputs at, for instance, a and b will activate the three counters a, b, and ab. Simultaneous inputs at a, b, and c will activate all the count-

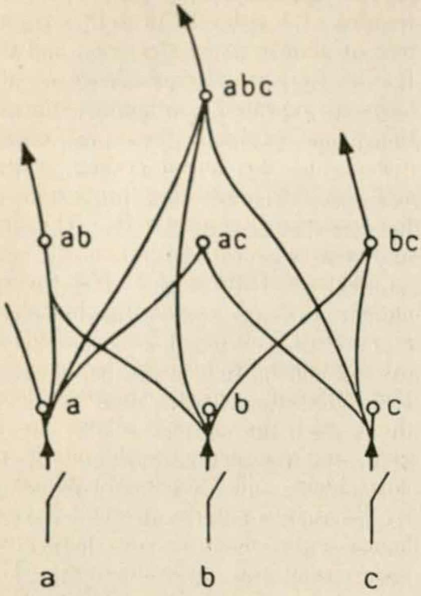


FIG. 2. A recognizer with three levels of classification. (For explanation, see text.)

ers. Suppose now that each counter has a limited memory, such that each time it counts, its state is altered stepwise, but gradually tends to revert to its initial condition. (A simple physical analogue is a leaky condenser, to which a standard charge is applied, each time the unit counts.) Suppose, further, that in this system two counters in approximately the same state (with equal charges) will count if either one of them is activated. For example, if *b* has almost always occurred at the same time as *a,* then if *a* is activated (but not *b*) the *b* and *ab* counters will also be activated. By choosing a suitable criterion for "same state" such a system computes conditional probabilities at a specified level. (The probability that *b* occurs, given *a,* determines whether or not the *b* and *ab* counters register when *a* occurs alone.) The details of these computers, and the related conditional certainty computers, are given by Uttley (1956, 1958, 1959) and will not be repeated here, except to note that by building in suitable short- and long-term decay characteristics, rather good analogues of the main features of classical conditioning can be produced. In any case it is clear that such computers will "recognize" patterns, conjunctions of patterns, and conjunctions of patterns occurring in rapid succession. Such conjunctions of patterns must lead to "generalization" or "inference" by the machine, and this is all that is needed to complete the type of perceptual learning required for the coupled system of coder and recognizer. For, in its normal environment, an organism is in constant visual contact with stimuli which fluctuate in size, shape, locus of stimulation of the retina, and so on. How does it come to "organize" or "recognize" a stable visual world? The usual answers are either that it learns to perceive the regularities amongst stimuli and conjunctions and sequences of stimuli, or that such perceptions are built in—both answers which nicely beg the question at issue when couched in such general terms.

The coupled system offers a solution which is not so flagrantly question begging. If one imagines a particular object in the early environment of an organism, say a circular disk, then as the organism approaches the object, moves around, over or away from it, the size, shape, orientation, etc. of the object will change, but these changes will be orderly, and will give rise to orderly changes in the output from the coder, and hence to orderly sequences of counts in the units of the recognizer. Many of these sequences will occur in fairly rapid succession (as when the organism locomotes) and, eventually, will give rise to "inferential learning" when the conditional probabilities become high enough, and stable enough. (In the interests of reasonable conciseness these conditions have not been enumerated; Uttley (1956, 1958) discusses them.) That is, when one member of the sequence is counted by the recognizer, if that counter's state is sufficiently close to the states of some other counters in the system, these will also register. The recognizer will therefore build up a system of internal classification which is dependent on the organism's history, to the extent that regularities in the perceived environment will be matched by regular patterns of activity in the recognizer. Thus, one can envisage two types of stimulus equivalence, one due to the ways in which the coder processes incoming patterns, which gives rise to "low order equivalence" because outputs from the coder are identical or very similar, the other due to experience of regular sequences of signals from the coder, which builds up higher orders of equivalence within the recognizer. Thus, for example, in this system the equivalence of a pair of circular contours is explainable in terms of the nearly identical outputs they generate in the coder. On the other hand, the "equivalence" of different perspective views of a circular disk, which may generate quite different outputs from the coder, is explainable in terms of "perceptual learning," which means, in this context, the regular conjunction of sequences of different signals. Such higher-order equivalence can even be thought of—rather speculatively—as the basis for perceptual constancy. It may be noted that this system offers an explicit and quite precise solution to the problems which Hebb's cell assembly theory was designed to meet. Hebb's explanation of "superordinate activity" as a basis for equivalence is vague, and in fact circular (Hebb, 1949, p. 98). The present system incorporates both the feature of low-level "preset" recognition of simple visual elements, and the feature of learned equivalence on the basis of repeated contiguous stimulation, but explains these as consequences of the initial coding system and the self-organizing properties of the recognizer, respectively. The first of these two deals satisfactorily with an awkward feature of Hebb's theory, namely that of explaining just how a *primitive* feature—a line, or angle in his argument—is initially "equivalent" for different retinal stimulus locations, and the second allows us to give an unambiguous definition of what Hebb calls "superordinate activity," that is to say, the constant physiological event that mediates perceptual recognition and generalization. The gain in precision leads to a more powerful theory of perceptual learning, which means of course one that is more amenable to experimental verification.

The testable consequences of this theory stem mainly from the properties of the coding system, and these predictions have been made, and some of them tested, in other places (e.g., Dodwell 1957, 1960, 1961a). However, the combined system does allow one to make a specific statement about perceptual equivalence, perceptual confusion, and the limits of discriminability. By the definition of the system, any pair of similar outputs from the coder are "equivalent," at the lowest level, for the recognizer. In view of the inferential character of the recognizer's operation, it is clear that, after some perceptual learning has occurred, a pair of similar outputs from the coder will activate a set of higher-order counters in the recognizer. If the pair of outputs actually are consequents of the reception of two different shapes, then perceptual confusion has occurred. However, any one shape (or object) will, in general, be capable of generating a number of different outputs from the coder, as explained above, and it is the set of these possible outputs that defines final, learned, recognition. Only in a case where *all* the possible outputs generated by a pair of shapes are the same would one have to admit the logical necessity of equivalence by confusion. Otherwise, successive discrimination of differences can occur in the system with prolonged visual inspection and search, as the two sets of units corresponding to the two shapes are organized in the recognizer. It should now be clear, too, that the system can handle contextual effects.

SOME NEUROPHYSIOLOGICAL EVIDENCE

The model presented in the previous three sections was devised and in part published some time ago (Dodwell 1957, 1958), but at that time there was behavioral evidence bearing only on the form of the coding system in any direct way. The recognition system was elaborated to complete the model for perception in a logically satisfactory way, as indicated in the Introduction above. It did no violence to the facts, yet there was no solid evidence in its favor—it was speculative. However, quite recently some neurophysiological findings which bear directly on this system have been reported, findings which confirm quite strikingly that the main features of the proposed coding and recognition system are correct, although modification in detail may eventually be required.

Hubel and Wiesel (1959, 1960, 1962) in a number of publications have reported on their method of direct recording from individual neurons in the striate cortex of cat. Their principal early finding (1959) was that many of these neurons have well-defined retinal receptive fields, which have a *specific orientation,* in contrast to retinal receptive fields at the lateral geniculate body, which tend to be circular. Moreover, such cortical cells characteristically are maximally sensitive to a stimulus which consists of a rectilinear contour falling on the midline of the receptive field, and in its orientation. Small changes in the orientation of the line relative to the retina are accompanied by rapid decreases in sensitivity, and small changes of position on the retina give rise to sharp changes in the firing pattern of the cortical neuron, since typically the receptive field consists of a narrow excitatory band, surrounded by a wider inhibitory field, or vice versa. It is *contour* rather than *surface* that activates these neurons; blanketing a receptive field with a surface leaves the neuron inactive. These so-called "simple" fields can be mapped out with small spots of light, to define their excitatory and inhibitory areas, and knowledge of these areas is sufficient to interpret the response of the neuron to more complex stimuli, such as a rectilinear contour, in terms of summation of excitation and inhibition from different parts of the field.

Clearly the simple receptive field represents a stage of processing sensory input which is very similar to the stage postulated for the coder described above. The latter's principal feature is that contours generate specific signals according to their orientation and position on the two-dimensional array. It will be recalled that these signals are taken into a function of a single variable, t, in the final common cable, in such a way that information about contour orientation is preserved, but positional information is not. Contours thus become *equiva-*

lent, so long as they are in the same orientation, and this is represented in the model by the fact that such equivalent contours all activate the same unit in the recognizer. Interestingly enough, Hubel and Wiesel (1962) have identified cortical neurons having exactly this property also; namely, that they are activated only by rectilinear contours in specific *orientations*, but the *position* of that contour on the retina can vary considerably, in some cases over a field subtending an angle of up to 32° at the eye. This clearly represents a further stage in the coding process, beyond the simple fields. To be specific:

> The cell . . . failed to respond to round spots of light, whether small or large. By trial and error with many shapes of stimulus it was discovered that the cell's firing could be influenced by a horizontally oriented slit ⅛° wide and 3° long. Provided the slit was horizontal its exact positioning within the 3°-diameter field was not critical. When it was shone anywhere above the centre of the receptive field . . . an 'off' response was obtained; 'on' responses were evoked throughout the lower half, . . . one might have expected wider slits to give increasingly better responses owing to summation within the upper or lower part of the field, and that illumination of either half by itself might be the most effective stimulus of all. The result was just the opposite: responses fell off rapidly as the stimulus was widened beyond about ⅛°, and large rectangles covering the entire lower or upper halves of the receptive field were quite ineffective. . . . On the other hand, summation could easily be demonstrated in a horizontal direction, since a slit ⅛° wide but extending only across part of the field was less effective than a longer one covering the entire width. One might also have expected the orientation of the slit to be unimportant as long as the stimulus was wholly confined to the region above the horizontal line or the region below. On the contrary, the orientation was critical, since a tilt of even a few degrees from the horizontal markedly reduced the response, even though the slit did not cross the boundary separating the upper and lower halves of the field . . . for this cell the strict requirements for shape and orientation of the stimulus were in marked contrast to the relatively large leeway of the stimulus in its ordinate position on the retina [Hubel & Wiesel, 1962, pp. 114–115].

At the same time it must be stated that these authors have found quite a variety of receptive fields, although the two types described appear to be the main ones. They find no evidence for a preponderance of neurons with receptive fields in horizontal and vertical orientations. No evidence on this point is available for the rat, but there is anatomical evidence of dominance in the octopus (Young, 1960), and there is probably a genuine species difference between the cat and rat in this respect, since Sutherland (1963) has recently shown that cats have no difficulty in discriminating between mutually perpendicular contours set at 45° to the horizontal, in contrast to the marked difficulty of this discrimination for the rat and for the octopus.

THE RESPONSE SYSTEM

Nothing has been said so far about responses, or the way classification in the system is related to response probability. There are two different problems to be faced. The first is: How are responses related to stimulation during the acquisition of hierarchical classifications in the recognizer? The second is: How are responses related to stimulation for an already formed (stable) classification system or subsystem? It would go beyond the scope of this paper to try to deal with these questions in detail, but a brief outline of a possible system might be as follows:

At some level of the classification system "response generators" are connected to the counters in the recognizer. These response generators have the property that, if activated, they produce a particular response or class of responses. Initially, all counters at the level in question have an equal probability of activating any one of the response generators (only one of which can be activated at a given time). The probability of a particular response generator's being activated by a particular counter on subsequent trials is raised or lowered according to the history of successes or failures of that response to the stimulus in question; there is ample scope here for deployment of the operators of stochastic models for learning. The relation of response generators to counters would be a one-many relation, both within and between different levels of the classification system. In this way discrimination learning would be possible at differing levels of complexity and abstraction. Early learning would occur principally at the lowest levels of the classification system (responses would only be made consistently to simple aspects of the environment), where adult learning could be in terms of more general, complex or abstract features. While this component of the system is indicated only in the briefest way, it at least shows that the initial suggestion made in this paper, of hooking together a coding and perceptual learning system with a response generating system, is quite feasible, and should lead to a more realistic characterization of discrimination learning processes than has previously been possible. Finally, it may be noted that perceptual learning in the proposed system is by contiguity alone, but that reinforcement of instrumental responses would certainly affect the contingencies through which perceptual learning and organization occur.

SUMMARY

The study of shape discrimination and discrimination learning has progressed from two different, but not contradictory, points of view. Theories of shape discrimination have concentrated on the problem of coding or analyzing the properties of shape as a basis for explaining the behavior of organisms with simple visual systems in discriminatory training and transfer situations. This approach has neglected the importance of perceptual learning. Theories of discrimination learning, couched in probabilistic terms, have neglected problems of stimulus classification and stimulus equivalence. A theory of shape classification (coding) is described, together with a system for recognition of the outputs of the coder, which is self-organizing in terms of regularities in the perceived environment. That is to say, the system incorporates the principle of perceptual learning. A possible method of connecting this coupled system to a response generating system is also briefly outlined. Taken together, these three systems—coder, recognizer, and response generator—constitute a logically satisfactory model for shape discrimination and discrimination learning.

REFERENCES

AYER, A. J. *The problem of knowledge.* London: Penguin Books, 1957.

DEUTSCH, J. A. A theory of shape recognition. *Brit. J. Psychol.,* 1955, **46**, 30–37.

DEUTSCH, J. A. The plexiform zone and shape recognition in the Octopus. *Nature,* 1960, **185**, 443–446.

DODWELL, P. C. Shape recognition in rats. *Brit. J. Psychol.,* 1957, **48**, 221–229.

DODWELL, P. C. Visual discrimination of shape in animals and men. Unpublished doctoral dissertation, Bodleian Library, Oxford University, 1958.

DODWELL, P. C. Discrimination of small shapes by the rat. *Quart. J. exp. Psychol.,* 1960, **12**, 237–242.

DODWELL, P. C. Visual orientation preferences in the rat. *Quart. J. exp. Psychol.,* 1961, **13**, 40–47 (a).

DODWELL, P. C. Facts and theories of shape discrimination. *Nature,* 1961, **191**, 578–581 (b).

Dodwell, P. C. Coding and learning in shape discrimination. *Psychol. Rev.*, 1961, **68**, 373–382 (c).

Hebb, D. O. *The organization of behaviour.* New York: Wiley, 1949.

Hilali, S., & I. C. Whitfield. Responses of the trapezoid body to acoustic stimulation with pure tones. *J. Physiol.*, 1953, 122, 158.

Hubel, D. H., & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.*, 1959, **148**, 574–591.

Hubel, D. H., & Wiesel, T. N. Receptive fields of optic nerve fibres in the spider monkey. *J. Physiol.*, 1960, **154**, 572–580.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 1962, **160**, 106–154.

Sutherland, N. S. Visual discrimination of shape by Octopus. *Brit. J. Psychol.*, 1957, **48**, 55–71.

Sutherland, N. S. Visual discrimination of shape by *Octopus:* Circles and squares, circles and triangles. *Quart. J. exp. Psychol.*, 1959, **11**, 24–32.

Sutherland, N. S. The visual discrimination of shape by *Octopus:* Squares and rectangles. *J. comp. physiol. Psychol.*, 1960, **53**, 95–103.

Sutherland, N. S. Cat's ability to discriminate oblique rectangles. *Science*, 1963, **139**, 209–210.

Uhr, L. Pattern recognition computers as models for form perception. *Psychol. Bull.*, 1963, **60**, 40–73.

Uttley, A. M. The classification of signals in the nervous system. *EEG Clin. Neurophysiol.*, 1954, **6**, 479–494.

Uttley, A. M. A theory of the mechanism of learning based on the computation of conditional probabilities. *Proc. First Internat. Congress on Cybernetics*, Namur, 1956.

Uttley, A. M. Conditional probability computing in a nervous system. In: *The mechanization of thought processes*, Her Majesty's Stationery Office, London, 1958.

Uttley, A. M. The design of conditional probability computers. *Information and Control*, 1959, **2**, 1–24.

Young, J. Z. The visual system of *Octopus. Nature*, 1960, **186**, 836–839.

(Received February 27, 1963)

Psychological Review
1964, Vol. 71, No. 2, 160-166

THEORETICAL NOTE

FURTHER ANALYSIS OF GSR CONDITIONING:

A REPLY TO STEWART, STERN, WINOKUR, AND FREDMAN

H. D. KIMMEL[1]

University of Florida

A reply to Stewart et al, (1961), regarding the "true" conditioned GSR. In addition to arguing that their "true" CR may be merely a part of a larger response which has gradually become inhibited, a definition of conditioning is suggested which permits responses to be classed as CRs even when they are elicited by the CS prior to training. Data are presented to show that the latency of the initial GSR to the CS *does* change during training, contrary to their claim that it does not. After many conditioning trials, the initial GSR gradually is transformed into a highly attenuated and delayed CR. The major difference between this point of view and that of Stewart et al. is that we define the "phenomenon" of conditioning by means of experimental control procedures rather than insisting that a new response to the CS occur.

In a recent article in this journal, Stewart, Stern, Winokur, and Fredman (1961) have questioned the validity of a considerable body of research on GSR conditioning on the grounds that adequate criteria for the definition of "true" CRs were neither specified nor met by previous workers in this field. Their argument was based principally upon acceptance of a stimulus-substitution definition of classical conditioning and upon the data of their experiment ($N = 19$). According to these writers, a true conditioned GSR is *a response to the conditioned stimulus (CS) which is not elicited by the CS before it has been paired with the unconditioned stimulus (UCS)*. Since almost any suprathreshold stimulus initially elicits a GSR prior to conditioning, they propose to distinguish between these original responses to the CS and true CRs by means of a latency criterion. Paraphrased briefly, this criterion is that the true CR has a latency outside the range of latency of the UCRs to the CS.

Applying the criterion to the data collected in their experiment, in which a 7.5-second CS-UCS interval was used, they report that responses of this sort are reliably discriminable from the original response to the CS, that their latency range is greater than that of the original responses to the CS, and that the response frequency curves of the two types of response during training are different. They concluded,

> In summary, we have put forward reasons for thinking that so far work on GSR conditioning has dealt with the adaptation and recovery of unconditioned responses rather than conditioning of responses [p. 66].

The present paper questions this conclusion on two grounds. It will be shown first that their true CR may be, in fact, a segment of what they refer to as the UCR to the CS and, second, that their approach to defining CRs may be the result of arbitrary reasoning about the conditioned GSR and classical conditioning in general.

Actually, Stewart et al. touched on the possibility that the deflection which they have called the true CR is part of the initial response to the CS. In the discussion of their Figure 2, containing a sample record, they note,

> The sample showing an example of three responses occurring at one trial illustrates

---

[1] Some of the research described was done under NIH Grant M-6060.

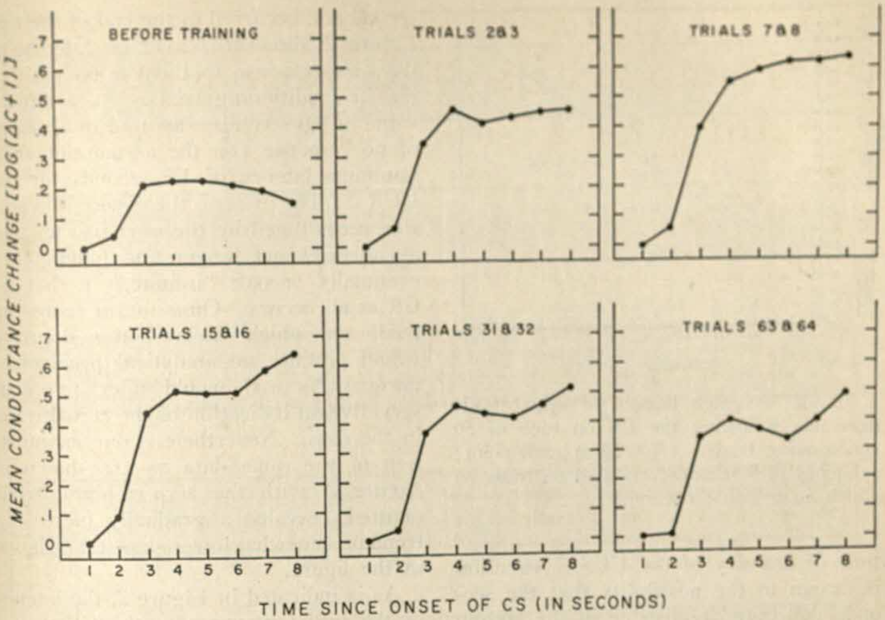


FIG. 1. Average magnitude patterns of the GSR during an 8-second CS–UCS interval before conditioning and at five stages of conditioning ($N = 8$ at each point).

> the fact that on occasion the conditioned response was *part of a complex of deflections rather than being discrete* [p. 64, italics added].

From the present point of view, double and other types of complex responses result from gradual changes in the pattern of the CR during conditioning, changes which parallel the phenomenon known as "inhibition of delay" in Pavlovian terminology. Inhibition of delay in GSR conditioning operates to depress the earlier portions of the initial CR (as well as to delay its onset) and frequently may create the impression of a "double response." Eventually (after many conditioning trials), only a very reduced and delayed response occurs, if there is any response at all. The present approach attempts to comprehend this change by observing trial-by-trial changes in the latency and pattern of the GSR to the CS.

In a recent experiment (Kimmel, 1963) a 1000-cps tone was used as the CS and an electric shock as the UCS. Eight subjects came to the laboratory at the same hour on 9 successive days and were given eight paired presentations of the CS and UCS on each day. The CS had a duration of 8 seconds and was followed immediately by a 1-second shock. The amount of difference between the basal resistance (pretrial) and the resistance at the peak of the first deflection of the recorder pen was measured at each 1-second point during the CS-UCS interval. This was done for each subject separately on each trial. These values were transformed into units of log conductance change and averaged for selected pairs of adjacent trials (three pairs of trials on the first day and the last two trials on the second, fourth, and eighth days). These data are shown in Figure 1.

Figure 1 shows that early in training the GSR reaches its peak magnitude after very few seconds and remains there during the entire CS-UCS interval. This is the reaction that Stewart et al. consider to be a sensitized original reaction to the CS. As training progresses, the strength of the early portion of this initial reaction declines, its latency increases, and its peak shifts toward the

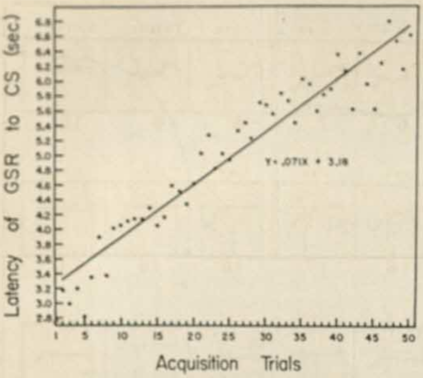


FIG. 2. Average latency of initial GSR deflection following the CS on each of 50 conditioning trials. ($N = 50$ at each point; a latency of 9 seconds was used on trials on which no response occurred.)

time of delivery of the UCS. Attention is drawn to the possibility that the second "deflection' noticeable in the bottom three panels of Figure 1 may not constitute a unique "second" reaction to the CS but, rather, may be what remains of the total reaction shown in the third panel (following 7–8 reinforcements), after the early and middle portions of the response have begun to be attenuated. Treating the early and late portions of the reaction as if they were different reactions, their frequency curves over trials would show the relationship reported by Stewart et al. (and in this sense replicate their findings).

A more elaborate study of the same phenomenon[2] indicates in greater detail exactly how continuous this shift in the latency of both the onset and the peak of the response to the CS actually is. In this study 50 subjects were conditioned at a 7.5-second CS-UCS interval, using a circular red light as the CS (produced by a Grayson-Stadler multiple stimulus projector) and an electric shock at the UCS. The shock had a duration of .1 second and occurred at the end of the CS. Figure 2 shows the average latency of the first response to the CS on each of the 50 conditioning trials. An arbitrary value of 9.0 seconds was used in all cases of no response (on the assumption of a minimum latency of 1.5 seconds for the UCR). The use of the 9-second value was necessitated by the obvious fact that as latencies get longer and longer they eventually become "infinite," in that no CR at all occurs. Omission of scores on trials on which no response occurred would produce an analytical problem of unequal *N*s and would offer too conservative an indication of the actual trend in the data. Nevertheless, our examination of the same data as are shown in Figure 2, with the zero-response trials omitted, revealed a gradually increasing trend of somewhat lower slope than shown in the figure.

As is indicated in Figure 2, the latency of the initial response to the CS moves upward in a continuous fashion from the first conditioning trials to the last. The straight line drawn through the data points was fitted by the method of least squares to focus attention upon the essentially linear nature of the trend. The slight drop in latency in the first few trials (e.g., Trial 5) may reflect an early sensitization effect, although sheer speculation is involved here. Consider-

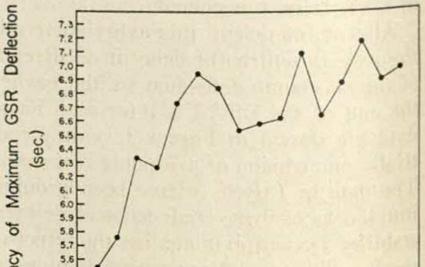


FIG. 3. Average latency of maximum GSR deflection following the CS on each of 16 blocks of three conditioning trials, omitting Trials 1 and 50. (Latencies for zero-response trials were omitted and *N*s reduced appropriately.)

---

[2] W. A. Greene, Ellen Kimmel, J. H. Kramer, and R. C. Martin collected the data in this experiment. Mrs. Kimmel's participation was supported by her National Institutes of Health predoctoral research fellowship PHS-MH-16798-02 (6S1A57).

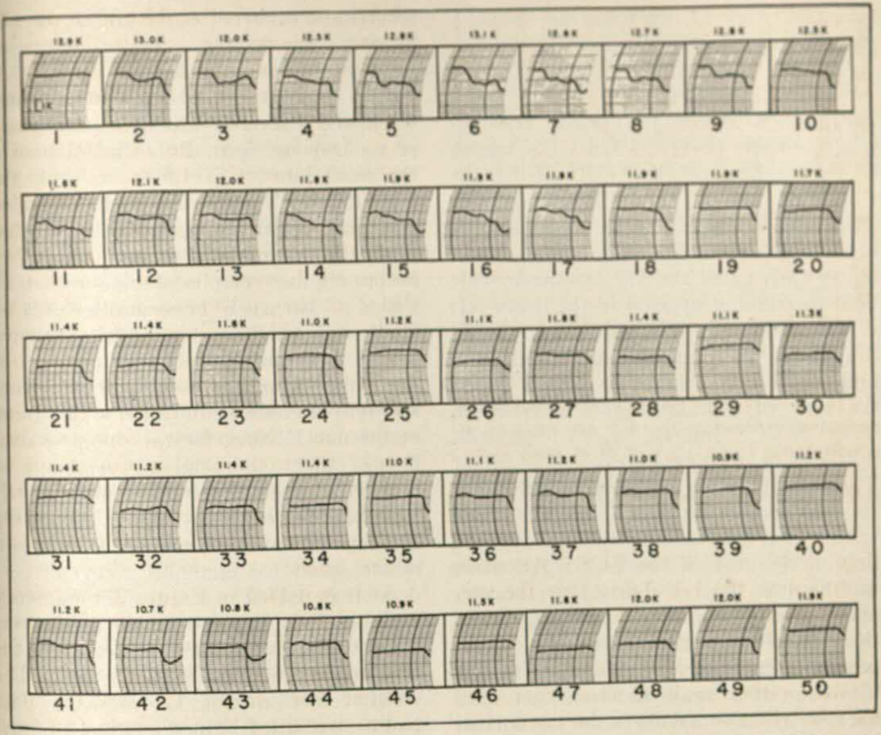


FIG. 4. Actual GSR record of a selected subject, who received 50 trials of CS–UCS with a 7.5-second interstimulus interval.

ing that each point in the figure represents the mean of 50 subjects, no statistical test appeared to be necessary to support the conclusion that a gradual increase in GSR latency occurred.

Also of interest in this experiment was the way in which the time of occurrence of the maximum deflection shifted toward the end of the CS-UCS interval. These data are shown in Figure 3, omitting all trials on which no response occurred. The data in Figure 3 have been grouped into blocks of three trials to increase their stability (to mitigate against the effect of small and unequal *N*s resulting from omitting trials of zero response). It is clear in Figure 3 that the increase in the average latency of the maximum GSR deflection approximates the form of a negative growth function rather than a straight line. This is to be expected for data of this type, since the latency of the maximum deflection, or "recruitment latency," as it is sometimes called (Prokasy, Hall, & Fawcett, 1962), reflects an excitatory as well as an inhibitory tendency (Dossett & Grings, 1959) (since the larger a response is, the later will be its maximum value).

On the basis of the foregoing data it would appear to be at least equally as reasonable to assert that the "second" deflection described by Stewart et al. is actually part of the dynamically changing *total CR* made to the CS as it would to argue that it is a unique reaction in its own right. As they have observed, two, three, four, and even more apparently separated deflections may occur following the CS during a long CS-UCS interval. In the present view, parsimony is sacrificed and controversy kindled by assuming that the second deflection is not only unique but the only true CR. Inspection of a selected actual GSR record, while not a proper basis for sweeping generalizations, serves to illustrate the difficulty that attends any effort to isolate the second deflection. Such an example is shown in Figure 4.

Figure 4 shows the actual trial-by-trial changes (over 50 CS-UCS trials) in the pattern of the GSR during the CS-UCS interval for a single subject. The basal resistance at the beginning of each trial is indicated (in ohms) at the top of each panel and the amplitude scale used throughout appears in the lower left of the first panel. The first heavy curved line shows the time of onset of the CS and the second heavy curved line shows the time of CS offset and UCS onset. The CS-UCS interval was 7.5 seconds.

It can be seen in Figure 4 that either a second deflection or a depression in the total deflection appeared on the third trial. On the fourth trial the second deflection is less discrete but larger, while on the fifth trial it has gotten smaller again and is followed by a third deflection. When we get to Trials 8 and 9 the second deflection has all but vanished. However, it is again noticeable on Trial 11. Trials 14 and 15 show a curious "stepladder" phenomenon, one which is not uncommon in our data. Finally, by Trial 19, both deflections are gone completely, although two very delayed CRs appeared on Trials 42 and 43. In the present view this record illustrates the *gradualness* and *variability* of CR changes during conditioning. It is because of this gradualness and variability that careful study of the trial-by-trial changes in the GSR pattern is required if we are to arrive at an understanding of the long-interval delayed conditioning process.

Stewart et al. (1961) are quite correct in observing, regarding the GSR, that:

> one finds that almost no stimuli are neutral, and hence that the reflex action proper to the conditioned stimulus is the same as that proper to the unconditioned stimulus, though varying in extent [p. 60].

What can be done, then, to eliminate the possibility that what appears to be "conditioning" may, by their definition, be mere "pseudoconditioning"? They would have us seek a GSR with latency characteristics different from those of the UCR to the CS, on the grounds that latency differences have been utilized by workers with other responses (e.g., eyeblink and startle reflexes). Although it was shown in Figure 2 that the latency of the conditioned GSR *does* increase during training (suggesting that a latency criterion could be used—i.e., the criterion of *increasing* latency), their position requires empirical support nevertheless. No such criterion is used in studies of salivary conditioning. Short-latency reactions in salivary conditioning are typically taken as indicating greater strength of conditioning than long-latency reactions (Kimble, 1961). Yet, Stewart et al. reject short-latency reactions as sensitized original responses and identify long-latency reactions as true CRs. The fact is that latency criteria have proved most useful, for example in eyelid conditioning, not as a means for describing what a CR *is*, but to identify what *is not* a CR. Here, the work of Grant (1945) on the *beta* response and of Spence and Ross (1959) on voluntary responses have had a significant influence in providing a basis for elimination of these classes of non-CRs from conditioning data. And, of course, this question cannot be dropped without noting that what may be appropriate for short-latency muscular reactions may not necessarily be appropriate for long-latency glandular reactions.

There is doubtless a need for differentiation among the various types of quasi-conditioning results and actual conditioning. One must start, obviously, with a definition of conditioning. Stewart et al. have chosen to apply to GSR conditioning a definition descriptively appropriate for salivary conditioning (stimulus substitution). It seems preferable to us to apply the label "conditioning" to any S-R alterations directly attributable to the *paired* presentation of the CS and UCS (Kimble, 1961) and "quasi-conditioning" to such changes not due to pairing. There appear to be only two acceptable procedures by which the conditioning–quasi-conditioning distinction can be made. The first would involve complete duplication, in a control group, of all the conditions characterizing the conditioning procedure, except the *pairing* of the CS and UCS. All that would then be required would be a significant difference between the conditioning and control groups (Kimmel, 1959). This procedure, of course, is the prototype for several of the more frequently used controls described in the conditioning literature, i.e., the "backward" control, the "random-unpaired" control, the "long interstimulus interval" control, the "pseudoconditioning" control, not all of which are equally adequate. The second procedure for differentiating between conditioning and quasi-conditioning would use differential reinforcement of at least two CSs. If after, say, eight paired presentations of a 1000-cps tone and an electric shock, interspersed with eight presentations of an 800-cps tone unpaired with shock, the GSR to the 1000-cps tone is significantly greater in magnitude than that to the 800-cps tone (Kimmel & Pennypacker, 1963), can we not attribute this difference to conditioning? According to the criterion of Stewart et al. that the CR not be elicited by the CS before training, the answer would have to be "No!" Both tones produced unconditioned GSRs before training. Nevertheless, if we succeeded in sensitizing the UCR to the 1000-cps tone, we would expect similar sensitization of the UCR to the 800-cps tone. Obviously, more than sensitization occurred.

The differential conditioning procedure serves very effectively to differentiate between *"specific* S-R sensitization," on the one hand, and sensitization due either to a lowering of the threshold of the particular response system for any stimulus or to an increase in general arousal, on the other, but does not differentiate between the latter two. The difference between GSR magnitude to the 1000- and 800-cps tones must have been due to either true conditioning or specific S-R sensitization. It could not have resulted from a general increase in GSR responsiveness (i.e., reduced GSR threshold) nor from an increase in general arousal (generally reduced thresholds), otherwise responses to *both* S+ and S− would have been affected. On the other hand, the random-unpaired control group technique, mentioned earlier, may isolate sensitization of the general arousal type. Kimmel (1959) ran three conditioning groups, each receiving a different intensity of the CS paired with electric shock, and three random-unpaired control groups, each receiving one of the three CS intensities used in the conditioning groups. The control groups received the same number of CSs and UCSs as did the conditioning groups in an unpaired, unsystematic sequence. The comparison between the conditioning and control groups receiving the weakest intensity of the CS (35 decibels) was the only one that showed a significant difference in favor of the conditioning group. A progressive increase during the training in GSR magnitude to the CS occurred in *both groups* that received the strongest intensity of the CS (115 decibels). This gradual increase in response magnitude was shown to be associated with a significant increase in basal skin conductance *from before to after training*, in both the conditioning and control groups receiving the 115-decibel stimulus (but not in the other groups), indicating that increased general arousal or activation underlay this instance of quasi-conditioning. It was surely true in this experiment that the relationship between CS intensity and conditioning was strikingly different from the relationship between CS intensity and quasi-conditioning, the former being negative but the latter positive.

SUMMARY

Having attempted to show that the true conditioned response identified by Stewart et al. may actually be a part of the earlier response to the CS which they reject as a sensitized UCR, we have chosen to define conditioning in such a way as to consider as CRs those reactions to the CS that may be qualitatively similar to responses elicited by the CS prior to training, providing that S-R changes are shown, by means of an effective control procedure, to be correlated with pairing of the CS and UCS.

It may be that, for the GSR, no effective approach can be found to differentiating between alpha conditioning, or sensitization (due to pairing) of a UCR to the CS, and true conditioning, in which the CS does not elicit the response before training (although latency changes were described that might do this). This implication of the present discussion should not be distressing since differentiation of this sort may not be necessary either. Kimble (1961) has recently arrived at a somewhat similar position with regard to the difference between eyelid conditioning and pseudoconditioning or, even, regarding "*all conditioning in which a noxious stimulus is employed* [p. 62, italics in original]."

The position taken in this paper is that Stewart et al. are correct in emphasizing the need for more adequate definition of the CR, but it is suggested that this be achieved via operational specification of the "phenomenon" of conditioning by means of effective control groups. Certainly those studies that have employed adequate quasi-conditioning controls can be taken to be studies of conditioning and not, merely, studies of "adaptation and recovery of unconditioned responses [Stewart et al., p. 66]." What is regrettable is that there are so few studies that meet even this criterion.

REFERENCES

DOSSETT, W. F., & GRINGS, W. W. A mathematical model for emotional arousal patterns. *Amer. Psychologist*, 1959, **14**, 425. (Abstract)

GRANT, D. A. A sensitized eyelid reaction related to the conditioned eyelid response. *J. exp. Psychol.*, 1945, **35**, 393–402.

KIMBLE, G. A. (Ed.). *Hilgard and Marquis' Conditioning and Learning.* (2nd ed.) New York: Appleton-Century-Crofts, 1961.

KIMMEL, H. D. Amount of conditioning and intensity of conditioned stimulus. *J. exp. Psychol.*, 1959, **58**, 283–288.

KIMMEL, H. D. Management of conditioned fear. *Psychol. Rep.*, 1963, **12**, 313–314.

KIMMEL, H. D., & PENNYPACKER, H. S. Differential GSR conditioning as a function of the CS–UCS interval. *J. exp. Psychol.*, 1963, **65**, 559–563.

PROKASY, W. F., FAWCETT, J. T., & HALL, J. F. Recruitment, latency, magnitude, and amplitude of the GSR as a function of interstimulus interval. *J. exp. Psychol.*, 1962, **64**, 513–518.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–385.

STEWART, M. A., STERN, J. A., WINOKUR, G., & FREDMAN, S. An analysis of GSR conditioning. *Psychol. Rev.*, 1961, **68**, 60–67.

(Received July 18, 1963)

VOL. 71, NO. 3 MAY 1964

PSYCHOLOGICAL REVIEW

TWO HYPOTHESES CONCERNING THE BIOCHEMICAL BASIS OF MEMORY

THOMAS K. LANDAUER[1]

Dartmouth College

The 1st hypothesis is that the basic event of conditioning or learning is the transfer of RNA molecules from surrounding glial cells into conducting neurons and the subsequent transformation of the protein synthesizing apparatus of the neuron. The transfer of RNA across membranes (of activated neurons only) is presumed to be due to electrophoretic migration in the reversed potential gradient during impulse conduction. The 2nd hypothesis is that information is coded for storage in the CNS in the form of frequency characteristics of spreading ac potentials to which a neural membrane can become tuned by alteration of its protein structure. Together, the 2 hypotheses suggest the basis of a physiological theory of learning.

Within the last few years considerable evidence has accumulated implicating ribonucleic acid (RNA), one of the macromolecules involved in the storage and utilization of genetic information, in the processes underlying memory storage in the nervous system. RNA is found in unusually great amounts in the grey areas of the central nervous system (CNS) (Mandel, Harth, & Borkowski, 1961). Its concentration in nervous tissue increases and decreases with age in correspondence with age changes in learning ability, and supplementation of RNA has been alleged to retard memory loss with age in humans (Cameron & Solyom, 1961). Supplementation of RNA has also been reported to improve learning and retention in rats (Cook, Davidson, Davis, Green, & Fellows, 1963), and chemicals which affect RNA utilization have been found to affect memory trace formation in planaria, mice, and rats (Chamberlain, Rothschild, & Gerard, 1963; Corning & John, 1961; Dingman & Sporn, 1961; Flexner, Flexner, & Stellar, 1963). The concentration of RNA in neural cells has been shown to increase as a result of stimulation (Hydén, 1959; Morrell, 1962), and this increase has been shown to be accompanied by a change in base ratio when the stimulation is part of a learning process (Hydén & Egyhâzi, 1962, 1963). Finally, Zelman, Kabat, Jacobson, and McConnell (1963) have reported preliminary evidence that injection of planaria with RNA extracted from conditioned planaria, as compared to RNA from untrained planaria, may facilitate conditioning.

---

[1] On leave at Center for Advanced Study in the Behavioral Sciences, Stanford, during 1963–64. I thank David S. Dennison, Joseph D. Harris, Wolfgang Köhler, Richard H. Rech, and William M. Smith for many helpful comments.

However, the manner in which RNA might enter into the learning process remains entirely problematical (Briggs & Kitto, 1962). Since RNA is intimately involved in protein synthesis, and thus in almost all possible structural modifications of cells, there is little limit to the number of kinds of neural mechanisms of learning which might reveal themselves in changes in RNA.

Suggestions to date include an unspecified mechanism involving direct coding of information on RNA molecules in terms of base sequences (Gaito, 1961), enzyme induction with respect to transmitter substances (Briggs & Kitto, 1962; Smith, 1962), changes in proteins of terminal vesicles (Hydén, 1961), and changes in ionic binding which influence firing rates (John, 1962). While these suggestions have much to recommend them, they do not by any means exhaust the possibilities for constructing a mechanism which accounts for the known facts. In the spirit of broadening the available spectrum of possibilities, a mechanism will be proposed here which differs in some fundamental ways from any suggested previously.

The critical features of the mechanism for which we are searching are that it provide a means by which RNA of selected neurons can be altered by experience, and a way in which altered RNA can give rise to altered behavior. Proposal of a mechanism thus constitutes the proposal of two interlocking hypotheses, a *learning* hypothesis, and a *retrieval* hypothesis. The two hypotheses to be presented here are suggested by consideration of some of the physiological events involved in classical conditioning. Together they suggest the rudiments of a theory of the processes underlying at least some forms of learning and memory.

HYPOTHESIS I: LEARNING

The first hypothesis, the *learning* hypothesis, is suggested in part by the fact that classical conditioning does not occur if the unconditioned stimulus (US) commences before or simultaneous with the conditioning stimulus (CS) (Kimble, 1961). The optimal condition is, rather, a .5-second interval between CS and US. Thus, rather than a mechanism by which the *simultaneous* activity of two neural elements results in a structural modification (e.g., Hebb, 1949; John, 1962), it may be more fruitful to look for a mechanism by which *sequential* action induces modification.

The neuronal events of the CS–US sequence, of course, consist of conduction of impulses and accompanying transient changes in membrane potentials, ionic concentrations, and membrane permeabilities throughout the system of cells involved. How might such changes produce alteration of the structures of neural elements?

A possible explanation would be available if impulse conduction resulted in the passage of new forms of RNA into neurons. If new RNA, which in some way represented the CS, were produced in the glia surrounding a neuron for a short period following the CS, but entered only neurons which then conducted impulses, new RNA would be incorporated by just those neurons activated by the US. If the result of incorporating new RNA which "represented" the CS (in a manner to be discussed below under Hypothesis II) were to alter the recipient cell so as to make it more likely to fire in the presence of the CS, the basic event of conditioning would have occurred.

Of course, macromolecules such as RNA do not readily cross cell membranes, although there are known cases in which they do (e.g., Mannick &

Egdahl, 1962; Schwartz & Rieke, 1962). However, there are dramatic changes in the membrane during impulse conduction which must tend to promote such migration. Specifically, the potential across the membrane (of the Squid axon) goes from approximately 60 millivolts negative inside to approximately 40 millivolts positive inside, the positive, or overshoot, portion of the impulse lasting about .5 millisecond (Tasaki, 1959). This represents a change from a resting potential gradient across the 50 A° thick membrane (Eccles, 1959) of 1.2×10^5 volts per centimeter to a transient potential gradient in the opposite direction which averages approximately 5.1×10^4 volts per centimeter (assuming the overshoot portion is approximately sinusoidal). That this change in forces should strongly promote the passage of RNA into the neuron is illustrated by the following calculations. The electrophoretic[2] mobility of yeast RNA (Chargaff & Zamenhof, 1948) at the pH of the brain (in a medium of phosphate buffer containing 0.15 M NaCl) is -12.1×10^{-5} cm^2 volts per second. This means that RNA in a gradient of 5.1×10^4 volts per centimeter, as present during impulse conduction, would migrate towards the positive pole of an electrical field at a rate of just over 6 centimeters per second (in the stated medium) and over the .5 millisecond of the reversed potential would travel approximately .003 centimeter, or some 6,000 times the thickness of the membrane. While there are a number of reasons to expect less migration across membranes *in vivo*, the size of the forces involved indicates that a considerable push toward the inside of the cell is given to any RNA surrounding it by the change in potential during conduction. Moreover, if the surrounding RNA has a "natural" tendency to enter neural cells (as, for example, the RNA of tobacco mosaic virus seems to have for plant cells, Fraenkel-Conrat, Singer, & Williams, 1957; and is apparently the case for several mouse tissues, Schwarz & Rieke, 1962; and amphibian neural tubes, Ficq, 1954), then an even more important factor would be the positive-outside resting potential which would act to shield the neuron against "infection" by new RNA *except* during conduction.

Thus, the proposed hypothesis is as follows. Glial cells are supposed to be specialized to provide RNA which, if it enters a neuron, will produce an alteration such that the neuron will become more sensitive to the events going on in the brain at the time the RNA was produced. Ordinarily, a barrier consisting of a potential gradient and membrane impermeability prevents the passage of this new RNA into neurons. However, when a neuron fires, this barrier is replaced by a potential gradient which actively promotes migration of RNA into the neuron. Thus, conducting neurons, and only conducting neurons, incorporate the newly available RNA.

If, at the time of conduction, the events going on in the brain were only those associated with the US, the incorporated RNA would be expected to be no different from that already present, and no change in the RNA population of the neuron would result. However, if a new stimulus, a

---

[2] *Electrophoresis* refers to the migration in an electric field of molecules as a result of the "atmosphere" of small ions which surround them. The rate of migration expressed as centimeters per second per volt per centimeter of the potential gradient is known as the *electrophoretic mobility*. The direction towards anode (−) or towards cathode (+) and the rate of migration is a function of the pH (as well as other properties) of the medium of diffusion. For further details, see Bull, 1951, pp. 161–191.

CS, had been acting very recently, new forms of RNA would be produced and incorporated, and the population of RNA types in the conducting elements would be altered as a result. The new RNA might act on the cell in a manner analogous to that of the "transforming principle" of bacteriophage (Ephrussi-Taylor, 1957) (which, however, consists of DNA rather than RNA) or the infectious RNA of tobacco mosaic virus which can produce modification of the protein synthesizing apparatus of the cell such that it thereafter produces the types specified by the invading RNA (Fraenkel-Conrat et al., 1957; Williams, 1959).[3] This picture agrees with the findings of Hydén, and coworkers (Hydén & Egyhâzi, 1962, 1963; Hydén & Pigon, 1960) that RNA decreases in the glia and increases in neurons during activity; that an alteration in base ratios occurs only when the stimulation is part of a learning process; that the new RNA is unlike that "native" to the neuron involved; and most significantly, that RNA changes occur in the surrounding glia as well as in the neurons involved in learning.[4]

HYPOTHESIS II: RETRIEVAL

The second hypothesis—the *retrieval* hypothesis—proposes a solution to the problem of how an altered population of RNA molecules and a resulting transformation of the protein synthesizing apparatus of the neurons activated by a US might make these elements responsive to the CS.

It should be noted that Hypothesis I could provide a mechanism by which synaptic changes, such as those postulated by Hebb and others, might occur. For example, the successive firing of two neurons which abut each other might result in a change in the protein structure of the membranes or terminal vesicals of the synapse such that the sensitivity of the second to stimulation by the first might be increased. The process involved could, for instance, be one of enzyme induction, in somewhat the manner proposed by Briggs and Kitto (1962) and Smith (1962). The mechanism of incorporation by electrophoresis would provide the necessary selective mechanism by which synaptic resistance would be lowered only between successively fired adjoining neurons. Given this basis for neural "association," the phenomena of learning and memory might be accounted for by the same sorts of theoretical machinery already so ably elaborated by such neural circuit designers as Hebb (1949) and Pitts and McCulloch (1947). However, in line with the general goal of this paper—to suggest alternative explanations to those now current—the temptation to follow well-worn pathways will be eschewed. Rather, a more radical hypothesis will be proposed, which seems to provide for a number of the "sophisticated" properties of the brain on the basis of a relatively simple mechanism.

The hypothesis is suggested in part by the fact that it is actually possible to observe electrical events associated with performance of conditioned responses with a fair degree of regularity and detail (e.g., John & Killam, 1960). If conditioning were solely a matter of the alteration of synaptic connections between individual pairs of neurons, the probability of observing activity specific to any one reaction would be very small indeed, as elements involved only in a specific reaction would be at least as small a proportion of the total brain as the conditioned reaction was of the animal's behavioral repertoire. Of course, electrical "signals" in the higher centers of the central nervous system are not confined to the firing elements, but, since neurons in these regions are unmyelinated, spread readity into surrounding tissues. Thus, recording electrodes need not actually be on those neurons which are imagined to have synaptic relations with each other in order to record electrical events generated by the elaboration of a conditioned reaction. However, it is hard to believe that this spread of electrical activity, and the perhaps specialized unmyelinization which allows it, serve no other function than to make the electrophysiologist's job easier and save a little space inside the skull case.

---

[3] "Transformation" has also been shown to be produced by infectious RNA extracted from animal viruses (Colter, Bird, Moyer, & Brown, 1957) and by RNA extracted from lymph node cells of rabbits (Mannick & Egdahl, 1962). In the experiment of Mannick and Egdahl, RNA from immunized lymph node cells altered recipient "neutral" lymph node cells to a state of transplantation immunity. Thus, precedent exists for the transformation of a mammalian cell by RNA from a cell of the same type but different function.

[4] The report of this last finding appeared after the present paper had been submitted for publication.

The suggested alternative is that conditioning consists not of decreased resistance of the synapse between one neuron and another, but rather in a change of the sensitivity of a given neuron to some characteristic of the spreading electrical activity generated by the CS.

It is proposed that individual neural cells selectively propagate and are excited by alternating currents of specific frequencies. As will be discussed below, neural membranes have certain electrical properties which could serve this function. In turn, the electrical properties of the membrane (or its "ghost" ionic envelope) are very likely a function of the structure of its protein layers, and these are in turn a function of the species of RNA involved in their synthesis. Thus, frequency specific sensitivity of the neural cell membrane provides an attractive possibility as a mechanism for signal selection based on biochemical structure.

Imagine that the membrane of each neuron (and its surrounding glia) constitutes an electronic filter tuned to some reasonably narrow band of frequencies. Then with conduction of an impulse the currents which spread from the neuron into surrounding tissues might have characteristic frequency patterns, probably represented in the case of a single impulse as the shape of the spike potential. These spreading potentials would then penetrate other neurons selectively, as determined by the frequency specific admittance characteristics (tuning) of their membranes. If a given neuron were tuned to just that band of frequencies which was present, the result would be at least partial depolarization of its membrane. Even if this depolarization were insufficient to fire the neuron, it might bias it to fire when provided with additional sources of excitation, e.g., synaptic chemical excitation from other neurons and/or potentials arising from intrinsic brain rhythms.

If the tuning of a given membrane is determined by the types of proteins of which it is constructed, then passage of transforming RNA from glia to a neuron might result in a change in its resonant frequency. If this change increased the probability of the neuron firing when exposed to frequencies characteristic of the CS, conditioning would have occurred.

Since each neural cell seems to have a unique biochemical composition (Sperry, 1951), it is conceivable that the membrane of each cell has a unique natural frequency or set of frequencies (a particular shape of pulse) and thus, in effect, is able to broadcast its genetic "code name." If a neuron which initially was unresponsive to the signal generated by firing of the neurons activated by a CS were "tuned" to respond to it, the functions of that neuron would come under the control of the CS. But, by Hypothesis I, the only neurons which can become so tuned are those which fire shortly after the CS, i.e., those initially under control of the US. Thus the *CS* would become a substitute for the *US* in eliciting neural activity.

INTERLOCKING OF HYPOTHESES I AND II

For the next step in the process, it becomes necessary to simply *assume* that the presence of particular FM signals results in the release of just those species of RNA which, if incorporated, cause a neuron to form membranes sensitive to the same frequencies. This, of course, leaves an important problem unexplained, since it is crucial to the theory that a mechanism for such RNA specification exist. How can a given frequency of electrical activity result in availability of RNA which breeds membranes sensitive to that frequency? Hydén (1962) has proposed that a direct "instructional" specification of new nucleotide arrangements is induced by ionic fluxes, although Briggs and Kitto (1962) and Smith (1962) have objected that there is no chemical precedent for such a process.[5] The alternative of a "selective" process is appealing in the present context, since it can be imagined that the glia are capable of providing a wide assortment of types of RNA unlike those of the neuron (their average base ratios are, indeed, different according to Hydén and Egyhâzi, 1962). Selection among these might proceed through stimulation of synthesis of resonating molecules, destructive selection against others, or selective migration, and/or membrane permeability. In any case, *if* appropriate RNA types were released by glia during, or shortly after the passage of a frequency coded signal representing the CS, and subsequently were absorbed by those neurons which fired, the appropriate neurons might be transformed so as to become more likely to fire in the presence of the CS.

In summary, the elaboration of a CS–US association is hypothesized to occur as follows. A CS causes a unique pattern of impulses in neurons of the cortical and subcortical projections of the afferents which it excites. Since each of these projection neurons has a unique biochemical "name," each propagates a spreading potential with characteristic ac frequency components. The compound FM signal thus produced is distributed widely in the brain, particularly in the "association" areas of the frontal and temporal poles, following the paths of least electrical resistance, e.g., flowing through the neuroglia. Glial cells throughout the regions reached by these signals are then, somehow, induced to make available RNA types which can produce membranes which respond to the same frequencies. These RNA molecules are presumed to accumulate at the glial-neural junctures during the CS–US interval. A US occurring at this point in time causes a certain number of neurons to fire, and the reversed electrical field of the action potential results in the electrophoretic migration of new RNA into these neurons. This RNA is then assumed to "transform" the protein synthesizing apparatus of the incorporating neurons in such a way that their membranes become sensitive to the FM signals which represent the CS. Thus, on subsequent occurrences of the CS these neurons are more likely to fire.

---

[5] However, what is in doubt is that *ionic fluxes* can direct differential RNA synthesis, not that differential RNA synthesis occurs. Cells differentiate during development despite identical genetic (DNA) composition; the actual types of RNA synthesized at any one time being a minute fraction of those possible with the given DNA. The process by which the synthesis of particular types of RNA is turned on and off is a deep enough mystery at present that, it seems to me, no conjecture ought to be considered unreasonable simply for lack of precedent. It seems at least conceivable, for example, that some sort of principle involving electrical and/or mechanical resonance of molecules which block parts of DNA chains could be involved.

SOME IMPLICATIONS AND RAMIFICATIONS

It may be useful to sketch the nature of a few major implications which these hypotheses would have for overall brain theory. One is that the function of the brain would depend not only on the neurons, but also on the neuroglial cells, as Galambos (1961) has proposed. Galambos' reasons for suspecting glia of more important functions than mere metabolic and mechanical support include the numerical preponderance of glial cells in the brain and the phylogenetic increase in the ratio of glia to neurons. The role proposed here for glia would be to provide preformed "programs," to use Galambos' language, bearing instructions to the absorbing neuron to respond to events representing certain cerebral inputs.

A second broad implication is that "association areas" need not be organized in a highly discrete manner. The envisaged process would occur most effectively in the densely packed, unmyelinated jungle of apical dendrites of the cerebral cortex where spreading impulses can reach large numbers of cells. The process involves essentially ephatic (or, following Galambos, "gliatic") rather than synaptic changes, and thus might be considerably restricted by myelinization. Thus, also, "conditioning" should be possible in any part of the nervous systems which contains densely packed unmyelinated neurons surrounded by glia (Chamberlain et al., 1963). The particular role of projection cortex implied by the theory is contained in its exquisite display in anatomical order of impulses arising from sensory stimulation. It transduces differences along stimulus dimensions to differences in anatomical location. The anatomical order, in turn, probably must be based on a biochemical differentiation, since it is retained in regeneration (Attardi & Sperry, 1963) and thus, on the present hypothesis, gives rise to differences in spreading signals which are characteristic of sensory inputs. For association all that is required structurally is that elements involved in response have rich connections with areas which also have rich connections with sensory projections. No more specific "pathway" organization is required, since "pathways" need consist only of conducting media which will allow the spread of FM signals.

The picture of neural interaction which is implied is somewhat different from traditional doctrine limited to neuron-neuron synapses mediated by chemical transmitters. However, it is not as radical a departure as it might seem at first glance. The major difference lies in utilization of the vast array of membrane—membrane juxtapositions found within the grey areas of the CNS which are not composed of two neurons, i.e., the neuron-glia and glia-glia interfaces. In a sense this view is more consistent with modern neuro-histology as disclosed by the electron microscope (Estable, 1961) than are traditional concepts based primarily on neural organization observed in the peripheral nerves and spinal cord. It has become clear that there is little actual intercellular space in the brain, but rather that every cell membrane abuts another. Apparently (Estable, 1961), only the neuron-neuron junctures are accompanied by terminal vesicles, and one might guess that these are responsible for one-way conduction of dc signals—all-or-none impulses—where the neuron-glia and glia-glia junctures which lack vesicles might be capable of passing ac signals only. The membranes involved do have excellent rectifying properties and high capacitances (Cole, 1941), from which such features might be expected. Moreover, the glia have very much higher membrane conductivities than do neurons (Hild, 1962), so that they might easily fulfill a second (or third, counting metabolic support) role which the present theory implies for them—the transmission of ac signals from one neuron to another over relatively large distances.

In addition to a storage mechanism, certain interesting possibilities for information processing features of the brain seem to be inherent in the hypotheses. For example, complex pattern qualities might be preserved in complex wave forms produced by interaction of signals from many separate sources, to which a given neuron or group of neurons could become conditioned. Generalization to nearby frequencies, and to harmonic and beat frequencies could be expected, and might form the basis of analysis or "categorization" of diverse inputs. Again, unspecialized cells which fired more or less randomly after sensory events or before responses could come to serve as relay stations, which might mediate between stimuli and responses and thus act as percepts or even "intervening variables."

Additional complexity and flexibility would be provided by the classical neuron-neuron synaptic mechanisms. While tuning provides a recognition or coding mechanism, it does not by itself provide adequate combinatorial or logical mechanisms. For example, it does not account for inhibition. It could be hypothesized that inborn, perhaps unmodifiable neuron-neuron synapses of the classical varieties make possible such features of neural activity. For example, tuning of a neuron which is genetically specialized as an inhibitor of a given response would produce an inhibitory CS. In a similar manner the conditioning of combinations of neurons with "naturally" convergent paths to different stimuli would allow synaptic summations to produce "novel," responses. If it is imagined that the regions in which conditioning occurs contain a wide variety of neurons, some excitatory, others inhibitory, some with many connections, others with few, then the combinatorial logic for information processing made available by allowing each to be selectively responsive to particular coded inputs (including signals generated by interneurons) becomes almost unlimited.

With elaboration, the hypotheses might also suggest new interpretations of such phenomena as consolidation, operant reinforcement, and discrimination—in fact, of course, of all the functions attributable to plastic properties of the nervous system—but an attempt to deal with all of these here obviously would be premature. Consequently, the foregoing sample of implications will be assumed to suffice to indicate some of the theoretical directions in which these hypotheses might lead.

Some Verifications

The ideas advanced here are testable on two different levels: directly, by investigation of the cellular and molecular events which are hypothesized, and indirectly, by study of the physiological and behavioral consequences which they imply. A few instances will be outlined in which evidence already exists.

Direct Evidence

The findings of Hydén and his collaborators (Hydén & Egyhâzi, 1962, 1963), which have already been mentioned, indicate that learning involves a change in the RNA of neurons and their surrounding glia, such that the glial cells lose RNA, the neurons gain RNA, and the variety of RNA in the glia and neurons is changed in a similar way. In rats which learned to climb a string to a food platform, increases in RNA were found in neurons of the vestibular nucleus and other directly connected neurons, but not in other parts of the brain. This increase in neuronal RNA, and an accompanying decrease in glial RNA was also found in control rats which received passive vestibular stimulation by centrifugation. However, changes in the base ratio (coding) of the RNA were found only in the vestibular neurons and glia of trained rats. All of these findings would be expected if a given type of RNA molecule were produced in glia and then transferred to the neuron. However, to prove that actual migration across the membrane during conduction is responsible for these changes, will require further careful investigation, perhaps by use of tracer experiments.

The question as to whether the RNA change is associated with a change in membrane constitution is also unanswered. (Although there is now evidence that *some* kind of protein modification is involved in learning, Flexner et al., 1963.) Microanalysis of protein types, e.g., lipoproteins, which occur primarily in the membrane by electrophoretic or immunochemical techniques, before and after learning, would be one way of searching for such changes.

There does not seem to be any direct evidence available relevant to Hypothesis II. However, it should not be impossible to detect selective sensitivity of cortical elements to specific frequencies of alternating currents or to detect differences in the frequency components of individual neural discharges. One guess as to a possible mechanism which might produce frequency specific membrane responses can be based on known electrical properties of neural membranes. Cole (1941) gives an equivalent circuit, based on ac bridge measurements, for the squid axon in which the membrane acts like a capacitor in parallel with a series resistor, inductance, and rectifier. Such a circuit has a natural or resonant frequency $f_0 = 1/(2\pi \sqrt{LC})$. When the resistance is less than $\frac{1}{2}\sqrt{LC}$, the current through the circuit oscillates at f_0, and when the resistance is much greater than this value, the circuit is overdamped at f_0. The rectifying properties of the membrane are such that resistance to inward current is approximately one hundred times that for outward (which, by the way, would prevent dc interneuronal transmission without chemical mediation). Thus, with the proper choice of values for each of the elements it would be possible to construct a circuit which would act either as an oscillator (when a dc potential was applied) or as a filter for ac currents. Outward flowing pulses would be given a particular shape, while arriving pulses of the same shape would generate the largest voltages across the circuit (and therefore produce depolarization contributing to a self-sustaining neural impulse). Thus, neural membranes are known to have properties which could make the broadcast of the "name" of a cell and the recognition of that "name" by a second cell—and thus Hypothesis II—possible. However, there is no evidence that neural cells do in fact function in this fashion or that such properties are related to learning. While frequencies of stimulating currents which produce minimums in threshold values have been observed for peripheral nerve and muscle fibers, and for central stimulation of certain kinds (Curtis & Cole, 1950; Rashevsky, 1950), the optimum frequencies in these cases being in the range of about 10 to 1,000 cps, differential optima for different neurons or regions do not seem to have been reported. It is conceivable that the morphologically specialized neurons of the cortex have membranes with extremely high resonant frequencies, perhaps in the high audio or RF ranges, due to very low inductance values. (It has even been suggested that the molecular structure of neural membranes constitutes a solid state device capable of propagating coherent electromagnetic energy in the VHF or IR range; Fernandez-Moran, 1962.) If such were the case, it would not be too surprising that frequency tuning had gone unnoticed. Nonetheless, in the present state of knowledge, a frequency specific tuning mechanism in the brain must remain purely hypothetical.

Indirect Evidence

A number of well-known but hard to explain phenomena of brain psychophysiology seem comprehensible on the basis of the present hypotheses. For example, the small size of the proposed memory record—one or a few cells—and the nonspecificity of its pathways would give the system the extensive insensitivity to damage classically described by Lashley (1960). Large areas might have to be removed before any of the few cells involved in particular habit were affected, and "connections" would not necessarily be severed by a scalpel's slice (or even by a gold or mica sheet which could be penetrated or by-passed by an ac field). The small area used for representation of an association would also give the brain the immense storage capacity required for multiple representation, which would provide additional indifference to insult.

One of the most striking features of brain function is the very marked temporal ordering of memory, which is seen not only in the facts of classical conditioning already mentioned, but also in the predominance of forward associations in human verbal learning (e.g., Raskin & Cook, 1937), and in such common experiences as inability to recall the previous letter in an alphabetical sequence (what comes before p?) without starting at an earlier letter and reciting forward. Just such a linear quality of memory, of course, follows from Hypothesis I.

On a more "molecular" level it is interesting to consider the relation of the proposed mechanism to the recent findings of McConnell (1962) and Corning and John (1961) of what appears to be purely biochemical transfer of memories in planaria. Following a report by McConnell, Jacobson, and Maynard (1959) that planaria retain conditioning after being transected and allowed to regenerate, Corning and John (1961) showed that only the head end retained the habit if regeneration took place in a medium containing ribonuclease, an enzyme which decomposes RNA. More recently McConnell (1962) has reported that digestion by untrained planaria of trained planaria resulted in savings in conditioning, and Zelman et al. (1963) have obtained evidence suggesting similar results from injection of an RNA extract of conditioned animals. It is obviously impossible to state, on the basis of the present theory, exactly how such a phenomenon would occur. However, one speculation might be that allowing altered cells to direct resynthesis of lost cells, as in regeneration, or feeding altered cells to another organism as in the "cannibalism" studies might "infect" the new animal with RNA which would make its neurons more sensitive to the CS. Only generalized "sensitization" would be expected, but this is not too implausible an explanation since the response was a very generalized contraction involving a large portion of the animal's effectors. Insofar as savings on relearning rather than direct transfer occurred in the cannibalism experiment, the diet of sophisticated worms may have provided only extra CS-coded RNA, which had to be incorporated by firing US-controlled neurons before it became functional for its consumers.

Finally, if the specification of signal (CS) sensitivity is carried by more-or-less freely migrating RNA molecules, a blood-brain barrier makes a good deal of adaptive sense. Not only could there be a need to prevent extraneous RNA from reaching neurons, but also there could be a need to keep the peculiar "coded" molecules of the brain out of the general circulation where they might induce immune reactions.

Summary and Discussion

Two hypotheses concerning a possible molecular basis of memory have been offered. The first is that the basic event of conditioning or learning is the transfer of RNA molecules from surrounding glial cells to conducting neurons and subsequent transformation of the protein synthesizing apparatus of the neuron. The second is that information is coded for storage in the CNS in the form of frequency characteristics of spreading ac potentials to which the membranes of neurons can be tuned.

The notion of a chemical interaction between glia and neurons is based in part on the recent findings of Hydén (Hydén, 1961; Hydén & Egyhâzi, 1962, 1963), and is at least vaguely in line with theoretical suggestions of Galambos (1961), while the idea of coding in terms of resonant frequencies is reminiscent of the theory of "modulation" and "resonance" once proposed by Weiss (1952) to account for developmental innervation. The suggested mechanism by which transfer of RNA occurs selectively to active (US–UR) neurons is electrophoretic migration in the potential gradient of the action potential. Transformation of the incorporating neuron by the RNA is assumed to be somewhat analagous to that in viral or bacteriophage infection (Williams, 1959). Selective tuning is suggested as possibly based on the electrical reactance and rectifying properties of the membrane as described by Cole (1941). No mechanism is suggested for the specification of the RNA type to be transferred. The lack of such a specification mechanism would, of course, represent a critical flaw in the theory, even in the event that all the other postulated mechanisms were correct. Principally to make this lack clear, the proposal has been stated in terms of two separate and "freestanding" hypotheses, instead of in the guise of a unified theory.

It is obvious that these proposals are highly speculative. There is no direct evidence that the brain uses the mechanisms suggested, and the indirect evidence is consistent rather than compelling. However, some of the ideas are quite different from current conceptions of brain mechanisms in learning, and it is hoped, therefore, that they will be of some interest.

REFERENCES

Attardi, D. G., & Sperry, R. W. Preferential selection of central pathways by regenerating optic fibers. *Exp. Neurol.*, 1963, **1**, 46–64.

Briggs, M. H., & Kitto, G. B. The molecular basis of memory and learning. *Psychol. Rev.*, 1962, **69**, 537–541.

Bull, H. B. *Physical biochemistry.* New York: Wiley, 1951.

Cameron, D. E., & Solyom, L. Effects of ribonucleic acid on memory. *Geriatrics*, 1961, **16**, 74–81.

Chamberlain, T. J., Rothschild, G. H., & Gerard, R. W. Drugs affecting RNA and learning. *Proc. Nat. Acad. Sci. U.S.*, 1963, **49**, 918–924.

Chargaff, E., & Zamenhof, S. The isolation of highly polymerized desoxypentose-nucleic acids from yeast cells. *J. biol. Chem.*, 1948, **173**, 327–355.

Cole, K. S. Rectification and inductance in the squid giant axon. *J. gen. Physiol.*, 1941, **24**, 771–788.

Colter, J. S., Bird, H. H., Moyer, W. W., & Brown, R. A. Infectivity of ribonucleic acid isolated from virus infected tissues. *Virology*, 1957, **4**, 522–532.

Cook, L., Davidson, A. B., Davis, D. J., Green, H., & Fellows, E. J. Ribonucleic acid: Effect on conditioned behavior in rats. *Science*, 1963, **141**, 268–269.

Corning, W. C., & John, E. R. Effect of ribonuclease on retention of conditioned response in regenerated planarians. *Science*, 1961, **134**, 1363–1365.

Curtis, H. J., & Cole, K. S. Nervous system: Excitation and propagation of nerve. In M. O. Glasser (Ed.), *Medical physics.* Vol. 2. Chicago: Year Book Publishers, 1950. Pp. 589–595.

Dingman, W., & Sporn, M. B. The incorporation of 8-azaguanine into rat brain RNA and its effect on maze-learning by the rat: An inquiry into the biochemical basis of memory. *J. psychiat. Res.*, 1961, **1**, 1–11.

Eccles, J. C. Neuron physiology—introduction. In J. Field & H. W. Magoun (Eds.), *Handbook of physiology-neurophysiology.* Vol. 1. Washington: American Physiological Society, 1959. Pp. 59–72.

Ephrussi-Taylor, H. X-ray inactivation studies on solutions of transforming DNA of pneumoccus. In W. D. McElroy & B. Glass (Eds.), *The chemical basis of heredity.* Baltimore: Johns Hopkins Press, 1957. Pp. 299–320.

Estable, C. Considerations on the histological bases of neurophysiology. In A. Fessard, R. W. Gerard, J. Konorski, & J. F. Delafresnaye (Eds.), *Brain mechanisms and learning.* Oxford: Blackwell, 1961.

Fernández-Morán, H. Molecular basis of specificity in membranes. In F. O. Schmitt (Ed.), *Macromolecular specificity and biological memory.* Cambridge: Massachusetts Institute Technology Press, 1962. Pp. 39–48.

Ficq, A. Analyse de l'induction neurale chez les amphibiens au moyen d'organisateurs marqués. *J. Embryol. exp. Morphol.*, 1954, **2**, 194–203.

Flexner, J. B., Flexner, L. B., & Stellar, E. Memory in mice as affected by intracerebral puromycin. *Science*, 1963, **141**, 57–59.

Fraenkel-Conrat, H., Singer, B. A., & Williams, R. C. The nature of the progeny of virus reconstituted from proteins and nucleic acid of different strains of tobacco mosaic virus. In W. D. McElroy & B. Glass (Eds.), *The chemical basis of heredity.* Baltimore: Johns Hopkins Press, 1957. Pp. 501–511.

Gaito, J. A biochemical approach to learning and memory. *Psychol. Rev.*, 1961, **68**, 288–292.

Galambos, R. A glia-neural theory of brain function. *Proc. Nat. Acad. Sci. U.S.*, 1961, **47**, 129–136.

Hebb, D. O. *Organization of behavior.* New York: Wiley, 1949.

Hild, W. Structure and function of neuroglia. In F. O. Schmitt (Ed.), *Macromolecular specificity and biological memory.* Cambridge: Massachusetts Institute Technology Press, 1962. Pp. 49–51.

Hydén, H. Biochemical changes in glial cells and nerve cells at varying activity. In *Biochemistry of the central nervous system, Proc. of the Fourth Int. Congr.*

of Biochemistry. Vol. 3. London: Pergamon Press, 1959. Pp. 64-89.

Hydén, H. Satellite cells in the nervous system. *Scient. American*, 1961, **205**, 62-70.

Hydén, H., & Egyhâzi, E. Nuclear RNA changes of nerve cells during a learning experiment in rats. *Proc. Nat. Acad. Sci. U.S.*, 1962, **48**, 1366–1373.

Hydén, H., & Egyhâzi, E. Glial RNA changes during a learning experiment in rats. *Proc. Nat. Acad. Sci. U.S.*, 1963, **49**, 618–624.

Hydén, H., & Pigon, A. A cytophysiological study of the functional relationship between oligodendroglial cells and nerve cells of Deiter's nucleus. *J. Neurochem.*, 1960, **6**, 57–72.

John, E. R. Studies of memory. In F. O. Schmitt (Ed.), *Macromolecular specificity and biological memory.* Cambridge: Massachusetts Institute Technology Press, 1962. Pp. 80–85.

John, E. R., & Killam, K. F. Electrophysiological correlates of differential approach-avoidance conditioning in cats. *J. nerv. ment. Dis.*, 1960, **131**, 183–201.

Kimble, G. A. *Hilgard and Marquis' "Conditioning and learning."* New York: Appleton-Century-Crofts, 1961.

Lashley, K. S. In search of the engram. In F. W. Beach, D. O. Hebb, C. T. Morgan, & H. W. Nissen (Eds.), *The neuropsychology of Lashley.* New York: McGraw-Hill, 1960. Pp. 478–505.

Mandel, P., Harth, S., & Borkowski, T. Metabolism of the nucleic acids in various zones of the brain. In S. S. Kety & J. Elkes (Eds.), *Regional neurochemistry, Proc. Fourth Int. Neurochem. Symp.*, 1960. London: Pergamon Press, 1961. Pp. 160–174.

Mannick, J. A., & Egdahl, R. H. Ribonucleic acid in "transformation" of lymphoid cells. *Science*, 1962, **137**, 976–977.

McConnell, J. V. Memory transfer through cannibalism in planaria. *J. Neuropsychiat.*, 1962, **3**, Suppl. No. 1, 42–48.

McConnell, J. V., Jacobsen, R., & Maynard, D. M. Apparent retention of a conditioned reflex following total regeneration in the planarian. *Amer. Psychologist*, 1959, **14**, 410. (Abstract)

Morrell, F. Electrochemical mechanisms and information storage in nerve cells. In F. O. Schmitt (Ed.), *Macromolecular specificity and biological memory.* Cambridge: Massachusetts Institute Technology Press, 1962. Pp. 73-79.

Pitts, W., & McCulloch, W. S. How we know universals. The perception of auditory and visual forms. *Bull. Math. Biophys.*, 1947, **9**, 127–147.

Rashevsky, N. Nervous system: Mathematical theory of its functions. In O. Glassner (Ed.), *Medical physics.* Vol. 2. Chicago: Year Book Publishers, 1950. Pp. 595–603.

Raskin, E., & Cook, S. W. The strength and direction of associations formed in the learning of nonsense syllables. *J. exp. Psychol.*, 1937, **20**, 381–395.

Schwarz, M. R., & Rieke, W. O. Appearance of radioactivity in mouse cells after administration of labeled macromolecular RNA. *Science*, 1962, **136**, 152–154.

Smith, C. E. Is memory a matter of enzyme induction? *Science*, 1962, **138**, 889.

Sperry, R. W. Mechanisms of neural maturation. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 236–280.

Tasaki, I. Conduction of the nerve impulse. In J. Field & H. W. Magoun (Eds.), *Handbook of physiology-neurophysiology.* Vol. 1. Washington: American Physiological Society, 1959. Pp. 75–121.

Weiss, P. Central versus peripheral factors in the development of coordination. In P. Bard (Ed.), Patterns of organization in the central nervous system, *Proc. Ass. Res. Ment. & Nerv. Diseases,* Vol. 30. Baltimore: Williams & Wilkins, 1952.

Williams, R. G. Replication of nucleic acids. In J. L. Oncley (Ed.), *Biophysical science—A study program.* New York: Wiley, 1959.

Zelman, A., Kabat, L., Jacobson, R., & McConnell, J. V. Transfer of training through injection of "conditioned" RNA into untrained planarians. *Worm Runners Digest*, 1963, **5**, 14–21.

(Received March 27, 1963)

Psychological Review
1964, Vol. 71, No. 3, 180-192

PERSONAL CONSTRUCTS, RULES, AND THE LOGIC OF CLINICAL ACTIVITY[1]

T. MISCHEL
Colgate University

The first part of this paper examines Kelly's *Psychology of Personal Constructs* and argues that constructs are not like hypotheses used to predict what will be done, but are like rules used to decide what should be done. The foregoing distinction is used to clarify the nature of personal constructs and their relation to behavior. The second part of the paper examines the logic of explanations and predictions based on such constructs. It argues that what is done by the personal construct clinician requires a "logical reconstruction" which differs sharply from the one provided by Meehl. The nature of explanations based on personal constructs, the way they can be empirically justified, and the relation of such explanations to predictions are analyzed.

ARE CONSTRUCTS PREDICTIVE "HYPOTHESES"?

The Fundamental Postulate of Kelly's (1959) theory is this: "A person's processes are psychologically channelized by the way in which he anticipates events [p. 46]." His theory thus explains behavior in terms of an "*anticipatory* rather than a *reactive* system [p. 170]." Kelly's fundamental point is:

> that human behavior may be viewed as basically anticipatory rather than reactive, and that new avenues of behavior open themselves to a person as he reconstrues the course of events surrounding him. Thus a thoughtful man is neither the prisoner of his environment nor the victim of his biography.
>
> The patterns of man's construction are called *constructs;* and since each person sets up his own network of pathways leading into the future, the concern of the psychologist is the study of personal constructs [p. 560].

These constructs are, allegedly, like hypotheses in that a man using constructs is like a "scientist, ever seeking to predict and control [p. 5] . . . the constructs one places upon events are working hypotheses, which are about to be put to the test of experience [p. 72]." Indeed, "constructs are hypotheses. Prediction is the goal [p. 940]."

Just what is a construct and how is it used to predict? Kelly (1955) holds that constructs are "ways of construing the world [p. 9]" and "by construing we mean 'placing an interpretation'. . . . In construing, the person notes features in a series of elements which characterize some of the elements and are particularly uncharacteristic of the others. Thus he erects constructs of similarity and contrast [p. 50]." Fear versus domination is one of Kelly's (1955) illustrations of a construct—a man who uses this construct divides the people he knows on the basis of this similarity–contrast dimension into those he is afraid of and those he can dominate (pp. 81–82).

Let us apply this to a particular instance of behavior; I insult my boss. Just how does this action involve a prediction? How is it "channelized" by the constructs I use to anticipate events? Kelly would say that if I construe people along the fear-domination axis, then I may perceive the boss as similar to some other people (e.g., my father), the basis of the similarity being that they tend to dominate others

---

[1] The author wishes to express his gratitude to the Colgate Research Council for a grant which enabled him to complete this study.

and are thus different from those who live in fear of others (the contrasting pole of my construct). I will, therefore, expect the boss to try to dominate me. If I construe my own role on the domination side of this construct, I may decide to insult him so as to show him that I cannot be dominated, I am someone he needs to fear. In this way my action has been "psychologically channelized" by the way I anticipate events. More generally, we can understand what people do in terms of their personal constructs because if people construe situations differently "they will anticipate them differently and will behave differently as a consequence of their anticipations [p. 90]."

The question I want to examine is whether there is a significant analogy between the process of construing which leads me to insult the boss and the use of hypotheses to make predictions.

1. The way I construe the situation may lead me to think "I will insult the boss because he will try to dominate me if I don't show him what's what." But this is not a prediction of what *I* will do any more than "I will take an aspirin because I have a headache" is a prediction. For it would be absurd to think that in such cases I *infer* what I am going to do from *evidence* which follows the "because." I am not offering the fact that I have a headache, or think the boss will dominate me, in support of a *prediction* which I am making about my own future behavior. In such contexts "I will" does not predict what I am going to do, but expresses my *intention* (decision) to do it; what follows the "because" is not evidence, but a *reason* which I give to justify what I am going to do as the appropriate thing to do.

2. Does the way I construe the boss lead me to predict what *he* will do? If I construe the boss as a dominating type I will, of course, expect him to behave in dominating ways. But this does not suffice to show that I am making a prediction. Pollyanna "expects" every cloud to have a silver lining, Mr. Smith "expects" to be home for dinner, the mad Hatter in *Alice* "expects" treacle to be drawn out of treacle well, and the physicist "expects" freely falling bodies to be uniformly accelerated —but they are doing different things and nothing but confusion can result from lumping their activities on the ground that all of them "expect" something. If the analogy between constructs and predictive hypotheses is not to be highly misleading, then there must really be a significant similarity between what I do when, on the basis of my constructs, I anticipate certain behaviors from another and what a scientist does when, on the basis of a hypothesis, he makes a prediction.

The following considerations bring out some radical differences between these activities.

(*a*) My anticipation of the boss' behavior tends to be a self-fulfilling prophecy. Because I construe him as dominating I insult him, and this is likely to make him do just what I "predicted" he would do when I construed him as dominating. But scientific predictions have no tendency towards self-fulfillment; what I do in predicting an eclipse in no way tends to make it happen.

(*b*) The validation or invalidation of anticipations based on constructs is "subjectively construed." Thus

> a person may anticipate that he will fall down the stairs and break his leg. If his prediction turns out to be true, or at least if *it seems to him* that he has fallen down the stairs and broken his leg, he experiences validation [pp. 158–159, 13; author's italics].

This subjective validation of anticipations based on constructs differs sharply from the objective verification of predictions. Facts about which there is intersubjective (i.e., objective) agreement determine whether the eclipse I predicted did, or did not, occur. But suppose I have construed the boss as hostile, he asks me to work overtime, and I construe this as validation. Was the request in fact hostile? The answer depends on the boss' intention—his request was not hostile unless he intended (consciously or not) to attack me in making it—and thus involves not merely construing but *construing how he construes.* Now when I construed him as hostile I made a *decision* about the kind of person he is; I am, therefore, "set" to construe whatever he does as an attempt to dominate. Even a conventional "How are you?" is something I could construe as having hostile intent—e.g., he is trying to show up my weakness by making me admit that I have not been feeling well. That such "validation" lacks objectivity is clear. What is not clear is how there could ever be something the boss does that could not be subjectively construed as validation. Is even a subjective invalidation of constructs possible?

(*c*) Kelly claims it is because a man "can usually assess the outcome of his predictions at a different level of construction from that at which he originally makes them [p. 13]." What he means is that since "no construct ever stands entirely alone [p. 304]," it may happen that, though I originally construed the boss as dominating, I cannot construe what he proceeds to do as dominating because to do so would conflict with some *other* constructs I use. In other words, such a construal would not fit into the "interweaving of construed likenesses and differences [p. 304]" involved in my *system* of constructs.

But would this invalidate my construct? Here it must be remembered that a construct is a dichotomous category involving both similarity and difference, e.g., fear versus domination. And this construct would *not* be invalidated when what the boss proceeds to do cannot be subjectively construed as an attempt to dominate. For as Kelly himself points out, what a person is likely to do in such a case is "to reverse himself along the dimensional lines which have already been established [pp. 128–129, 133–134]." Though my "prediction" has been subjectively invalidated, I am likely to retain the construct on which it was based and to use it now to construe the boss as someone who is afraid. Suppose this construal in turn is subjectively invalidated by what he proceeds to do. Will this finally force me to reject the construct as inconsistent with the facts? No again. For in that case I can "constrict" the construct, render it less "permeable," and thus exclude the boss from its "range of convenience [pp. 477, 79–80, 68–69, 77, 137]." Thus it is not logically possible for the facts ever to falsify my construct—not even subjectively. Surely this is why an experimental study of the effects of interpersonal predictions on changes in personal constructs concludes that the

> findings with respect to the effect of invalidation on construct change are inconclusive . . . [there was a] failure of invalidation to show a statistically significant relationship to changes in constructs [Poch, 1952, pp. 85–86].

Since constructs cannot be falsified (not even subjectively) by what in fact happens, they differ fundamentally from predictive hypotheses which can be.

3. Another fundamental difference between construing and predicting becomes clear when we note that "hostile," "dominating," "mature," "gentle," and other poles of personal con-

structs are not only *descriptions* of behavior, they are also *evaluations*. My construal of the request to work overtime as hostile is not value-neutral; to construe the boss as dominating is not like perceiving something as an apple, but like perceiving it as a bad apple.

If construing did not involve evaluation it could not channelize behavior. For my prediction that the boss will try to dominate me can channelize my behavior towards him only insofar as I value or disvalue the anticipated outcome. If I don't care one way or the other about dominating or being dominated, then my anticipation that he will dominate me gives me no reason for acting one way rather than another. So if my behavior is to be psychologically channelized by the way I construe, then my construals must involve an evaluation of what I am construing. But a prediction, as such, never evaluates what it predicts.

PREDICTIONS AND HYPOTHESES VERSUS DECISIONS AND RULES

I have argued that using constructs is not like using hypotheses to make predictions. What leads Kelly to think it is? He claims that the girl who "anticipates" marriage to a man having certain characteristics, the old maid who "predicts" a husband that never comes along, and the navigator who predicts his arrival at the North Pole are all doing the same sort of thing—they are using constructs to predict the future. According to Kelly, what each of them predicts is the "common interesect of a certain set of properties," and the only difference between them is that in the case of the old maid "there are altogether too many dimensions involved and nobody ever lands on the precise point where all of them converge. Her long-standing anticipation is never fulfilled [p. 121]." This is why Kelly thinks that the behavior of the girl and the old maid is like that of the navigator—a matter of using "hypotheses" (constructs) to anticipate the future.

But this confuses entirely different things. Though the navigator may say "I ought to (will, should) see a place having such characteristics," and this sounds very much like the girl saying "I ought to (will, should) marry a man having such characteristics," what they mean is different. For the navigator's "ought" is *predictive* while the girl's "ought" is *prescriptive*.

Consider a typical prediction: "I should see an eclipse at such a time." Certain facts (i.e., planetary positions and dimensions), together with certain laws (i.e., Kepler's), are the basis for an inference about when the moon will pass into the earth's shadow to produce a lunar eclipse. We are predicting the time of the eclipse by predicting the future positions of the moon and the earth from their present ones, and the "connection" between these data is effected by laws formulating uniformities in planetary behavior. When we are not fully confident of the uniformity on which our prediction is based, we call it a hypothesis rather than a law and wait for the facts to confirm or refute it.

In *predicting*, I thus describe what will happen by drawing an inference on the basis of some *laws* (*hypotheses*). These laws *describe*, e.g., planetary behavior, but they do not evaluate it. They are not directions laid down to guide the planets towards the "right" motions; the planets did not "decide" to follow these laws, nor could they decide to disobey them. When Kepler first formulated his hypotheses and used them to predict planetary motions, this could have had no tendency to make happen what he predicted would happen. The facts, "objectively construed," determine whether his hypotheses are true or false.

Compare this with the girl who says "I should marry a man having such characteristics." She is not predicting what will happen, but is expressing her intention to marry that kind of man. She has *decided* that this is the "right" sort of man for her. Her decision is not an inference based on laws describing what in fact happens; it is a prescription for what she should do based on *rules* she follows in deciding what to do. These rules, unlike laws (hypotheses), guide and direct her activity by *evaluating* certain behaviors as the right (appropriate) things to do. (e.g., "I should be a dominating person, like father, not one who lives in fear.") She has decided to follow this rule and she could disobey it. In following the rule she makes things happen—i.e., she turns down this suitor and that because he does not have the requisite characteristics. The rule she follows thus tends to bring about what it prescribes; indeed, that is its function. What in fact happens cannot show the truth or falsity of the rules she follows. For rules prescribing what *should be done* cannot (logically) be confirmed or disconfirmed by the facts; unlike hypotheses which describe what *is done,* they cannot be true or false. But it can be foolish or self-destructive for someone to follow certain rules. The trouble with the old maid who never finds a husband is not that she lacks skill in using hypotheses to make predictions; it is that she follows foolish rules—the conditions she lays down for the right man involve "altogether too many dimensions" and so she decides to turn down every suitor.

Kelly's discussion of construing thus fails to distinguish between the way *rules* which prescribe behavior are used to *decide* what one should make happen, and the way *hypotheses* (laws) which describe behavior are used to *predict* what will happen.

All this is not to suggest that Kelly's approach is "unscientific," nor to advocate that clinicians model their work on the theories developed by physicists. I will deal with metatheoretical questions in the second part of this paper and will there defend the use of personal constructs for the explanation of human actions. My aim here is to clarify the nature and function of personal constructs. What I am arguing is that there are important logical differences between using hypotheses to make predictions and using rules to make decisions, and that construing resembles the latter rather than the former.

RELATION OF CONSTRUCTS TO BEHAVIOR

The foregoing distinctions can now be used to specify the way constructs channelize behavior. To use constructs is, first of all, to *follow rules* in classifying people. If I construe in terms of fear versus domination, I have decided that people should be classified as "similar" or "different" on the basis of whether they dominate others or live in fear of others. The fear–domination construct is not a hypothesis I use to predict that the people I meet will be either dominating or fearful types; I perceive the people I meet as dominating or fearful types because I classify them on the basis of the fear versus domination construct.

Finding that, e.g., my boss exhibits certain behaviors which fall under the domination pole of my construct, I classify him as a person who tries to dominate others. Since to classify something as an X is to expect it to behave in X-ish ways, this construal of the boss will lead me to expect dominating behavior from him. But the subjective anticipations concerning the future behavior of another which follow from my construal (classification) of him cannot, of themselves, channelize *my* behavior. For they provide me with no basis for action unless I also *evaluate* what I anticipate. Constructs provide such evaluation because, as Kelly (1955) puts it:

> The person builds his life upon one or the other of the alternatives represented in each of his dichotomies. This is to say that he places relative values upon the ends of his dichotomies. Some of the values are quite transient . . . others are quite stable and represent guiding principles. Even the stable ones are not necessarily highly intellectualized—they may appear, rather, as appetitive preferences [p. 65].

Because one end of the fear versus domination construct is "valued" over the other, the construct functions as a rule in terms of which I can decide what I *should do* about someone whom I have construed this way. The construct "governs [p. 132]" my behavior because it is a rule I use in deciding what the right thing for me *to do* is.

Thus my fear versus domination construct not only leads me to *classify* the boss with respect to this dimension, it also leads me to see this dimension as the one in terms of which *I must act:* Either I allow the boss to dominate me and live in fear of him, or I try to dominate him and make him fear me. If I value the domination pole, then the latter will be the right course for me to follow. On the basis of such a "calculation," I may decide to insult the boss (see the Choice Corollary, Kelly, 1955, p. 64; also pp. 487–488, 497–498, 521–526). If asked why I insulted the boss I will bring out my anticipations about what would happen if I did, or did not, insult him. These anticipations are my *reason* for insulting him. What makes them a reason to act this way rather than another is a *rule* I follow in deciding what the right thing to do is. It is this rule which "connects" the subjective anticipations following the "because" in "I will insult him because . . ." with my decision to insult him.

My behavior towards the boss has been psychologically channelized by the fear versus domination construct because my decision to construe him as dominating and the subjective anticipations which followed from it, as well as my decision to insult him in order to dominate him instead of having to fear him, have been guided by the construct—*the construct is the rule I followed in making these decisions.* But the discussion has been oversimplified since actual decisions are likely to be determined by the intersection of a number of construct dimensions rather than by a single construct. Since constructs are arranged in a system of "ordinal relationships"—some being like "fundamental principles" while others are "subordinate [pp. 56, 61, 78]"—the system of personal constructs is the hierarchical system of rules a person uses to guide and direct his actions. That is why personal constructs channelize a man's actions and can be used to explain them.

I now turn to metatheoretical considerations. Just how do personal constructs, understood in light of the foregoing discussion, serve to explain and predict an individual's behavior? Are there significant *logical* differences between such explanations and predictions on the one hand, and those given by S–R, or need or drive theorists, on the other?

As a point of departure, consider Meehl's (1954) "Logical Reconstruction of Clinical Activity" (Ch. 6). Meehl argues that:

> every skilled clinician must be making use of some laws, however vague, which may be of considerable generality, but which nevertheless make it possible for him to order his material with respect to a given patient in terms of some general nomothetic basic psychodynamics [p. 53].

In "forming the concept of this person" the clinician is "assessing his needs and his modes of satisfying those needs," and this means that he has to

> utilize the given facts, together with crudely formulated laws, to invent a hypothesis concerning the state of certain intervening variables or hypothetical constructs in his patient [p. 46].

The law the clinician uses is nomothetic in its form, but "in its *end terms,* i.e., in the defining properties . . . that specify 'S', 'R', 'G', 'D', etc., strongly idiographic [p. 64; also pp. 42–44, 60]." It is very difficult to specify the general defining properties of these classes "because in the case of human social behavior the defining properties are, in general, not physicalistic [p. 54]." Consequently, it is only in light of a hypothesis about the particular patient that one can see that a certain behavior datum belongs to, e.g., a certain S or R class, and thus can apply the general law to the particular case (pp. 52, 54–55, 82). Moreover, though this hypothesis about the state of the patient must, in conjunction with the law, logically entail the data, it cannot be logically deduced from the data (p. 57). The creation of such hypotheses about the states of variables in the particular patient is, according to Meehl, the unique contribution of the clinician; it is the reason why he cannot be replaced by a Hollerith machine (pp. 49–50, 65–66, 78).

In what follows I will argue that when a clinician explains and predicts his patient's behavior in terms of personal constructs, what he is doing requires a "logical reconstruction" that differs sharply from the one offered by Meehl.

CLINICAL EXPLANATIONS

What a clinician like Kelly does in "forming the concept of this person" cannot be understood as "assessing his needs and his modes of satisfying those needs." For Kelly (1955) explicitly rejects all "*push* theories based on 'stimuli' and *pull* theories based on 'needs' [p. 36, his italics; also 1960, pp. 49–50; 1962, pp. 84–86]." Against such theories Kelly (1960) insists that the clinician must "not envision the behavior of man in terms of the external forces bearing upon him [p. 60]"; he "must not construe his client as an unthinking object; he must construe his client's construings. This is the central theme of the psychology of personal constructs [1955, p. 531]." And this implies that there are important logical differences between explanations in terms of personal constructs and the type of explanation reconstructed by Meehl.

1. For Kelly the clinician must account "for the alternative a man selects in a choice situation [p. 45]" in terms of the way *that man himself construes* his situation. Where Meehl sees the clinician as hypothesizing about the states of "process variables" (McClelland, 1955) in order to apply nomothetic laws to the behavior of the patient, Kelly sees the clinician as eliciting the *content* of the agent's personal constructs—the rules in terms of which this person defined what alternatives he had to choose between and what the right thing to choose is. So while Meehl's account takes the point of view of an external spectator of people's behavior and emphasizes "psychological process" (i.e., the application of general principles about *how* people satisfy needs to the particular

case), personal construct psychology takes the point of view of an agent who makes choices and emphasizes "psychological content" (i.e., *what* constructs on the part of this person lead him to make such choices).

2. On Meehl's (1954) reconstruction, the clinician explains behavior by showing that it is predictable (i.e., entailed by laws in conjunction with an hypothesis about the state of the patient) and thus not surprising. Clinical explanation is clinical pre- or postdiction (p. 123). In contrast to this, the personal construct clinician uses the content of the patient's constructs to show that because the patient followed these rules he chose this as the right thing to do. Instead of showing how the behavior to be explained can be subsumed under general *laws,* the personal construct clinician shows that if one followed these *rules* (i.e., used these constructs) then this would seem the right thing to do. In one case we understand behavior by seeing its occurrence as *predictable*; in the other, by seeing its choice as *intelligible*.

3. Since explanations in terms of personal constructs are given from the agent's point of view, they must involve the sort of *concepts the agent himself understands and uses in deciding what to do.* Fear versus domination, gentle versus aggressive, etc., are considerations of the kind an agent uses in deciding what to do. This differs sharply from Meehl's clinician whose explanatory concepts—e.g., drive states, needs, reaction formations, etc. —are not the sort an agent uses to explain *his own* behavior. Whether the person is capable of understanding Meehl's kind of explanation is irrelevant to its correctness; but since personal constructs explain from the agent's point of view, the explanation cannot be correct if it involves concepts which the agent cannot grasp and thus cannot have used in deciding what to do. This does not mean that whatever explanation the agent gives for his own behavior is correct, or even that he can verbalize the constructs (rules) on which his decision was based. For the agent may not be aware of the content of some of his constructs, or of the way he uses them, because they are at a "low level of cognitive awareness"—i.e., some of the constructs involved may be "preverbal," or one end of a construct may be "submerged," or a construct may have been "suspended" (Kelly, 1955, pp. 459–476; also Mischel, 1963). In a clinical setting the patient, typically, is unable to *tell* correctly, or at all, how he really construes and it is the clinician's job to discover the rules on which he bases his choices. Nevertheless, the constructs used to explain behavior must be such that in principle the agent could use them to make his choices.

Explaining behavior in terms of personal constructs thus differs from explaining the behavior of an "unthinking object" in terms of "pushes or pulls"; to give such explanations the clinician must try to "understand the client by putting himself in the client's shoes and attempting to stand and walk like the client [1955, p. 333]."

JUSTIFICATION OF CLINICAL EXPLANATIONS

Many psychologists see taking the agent's point of view as a way of *discovering* a hypothesis that may explain the patient's behavior; but since the intuitive plausibility of a hypothesis is no guarantee of its truth, they often insist that the only way it can be *justified* is by showing that in conjunction with general (nomothetic) principles it entails such behavior. Thus Meehl (1954), although he is "impressed" by the "vacuous character [of general principles] insofar as generating the particular hypothesis is concerned [p. 50]," insists that a hypothesis about the patient can be justified only by showing that in conjunction with general principles it *entails* his behavior—this is the only "scientific," or even "rational," support that can be given for it (pp. 57, 66, 72–73, 81). Similarly, Falk (1956) holds that we can "turn to idiographic procedures for the provision of working hypotheses, not for their confirmation [p. 61]."

Yet when we turn from theoretical discussions to clinical practice there is something oddly irrelevant about the nomothetic principles which, allegedly, are needed to justify clinical explanations. Consider Meehl's (1954) explanation of a patient's dream in which she feels resentful towards someone dressed in a riding habit. He interprets the dream as involving a sexual pun on the phrase "riding habit" so that the person in the dream is her boy friend with whom she does not want to establish the habit of having sexual relations (p. 71). Is it really the

> use of such general principles as these: in dreams abstract notions are often represented by concrete forms and processes and in a minority of cases the mode of plastic representation involves a pun [p. 72]

which justify the clinician's hypothesis? Meehl himself says that "the remainder of the dream and her associations to it . . . confirmed this hypothesis [p. 71]"! What actually justifies the clinician's hypothesis seems to be the way his explanation "fits" (makes intelligible) other things the patient says and does; the general principles seem "vacuous" not only with respect to the discovery, but also with respect to the justification of such clinical hypotheses. (Similar considerations apply to Meehl's other illustrations; vide pp. 48–50.)

What leads writers like Meehl to insist that general laws ("however vague") *must* be used to justify clinical explanations is not the facts of clinical practice—the very "vague" and "vacuous" character of these laws indicates that they can be of no use to the practicing clinician. What really lies behind this insistence is a theoretical point about the *logic* of explanation. The model of explanation which most psychologists have in mind derives from the physical sciences, and in those sciences the "ordering" of the data, the "connection" between the fact being explained (e.g., the time of an eclipse) and the facts offered in explanation (e.g., planetary positions, dimensions, etc.) is based on general principles. Since the event to be explained and the events offered in explanation are separate and distinct, they can be connected, or ordered, only by means of general principles asserting a relation between such events. Now if we view a person's acts from the standpoint of an external spectator, then they will be seen as "behavior data," i.e., happenings to be explained on the basis of the foregoing model. The sexual pun in the dream will then be regarded as an "event" whose occurrence can only be explained by connecting it with other events. Since this can only be done by means of general principles, a psychologist who construes the situation in this way will claim that it is only in terms of (nomothetic) laws that one can (logically) "order his materials with respect to a given patient," or "rationally justify" one's explanations.

But the situation is radically different when we explain behavior from the agent's point of view, i.e., in terms of the way he construes. For the sexual pun in the dream will then be seen not merely as a *happening* but as an *action*—in dreaming this the patient is expressing her intention not to establish habitual sexual relations with her boy friend. Why she has that intention can be explained (i.e., connected with the rest of her behavior) by showing how the way she construes would make this the wrong thing for her to do. Why she expresses this intention in a dream, rather than overtly, can be explained by showing how the way she construes would make an overt expression of hostility wrong. Since constructs are the *rules* a person follows in deciding what *to do*, they are themselves the ordering principles which connect the action with the construal that explains it. Personal constructs explain behavior, not by showing that such behavior data are deducible from these constructs in conjunction with general principles, but by showing that such actions are intelligible because this would seem the right thing to do if one construed this way. Since such explanations are based on the content of the rules an agent follows in making choices, rather than on general laws connecting various kinds of happenings, they stand in no need of nomothetic laws for their justification. (For a fuller discussion of this issue, see Mischel, 1963).

But this does not mean that giving such explanations is an intuitive, nonempirical procedure. The clinician must *discover* how his client construes from the facts and he must *justify* his explanation in terms of the facts—i.e. what this person says and does, or would say and do under certain circumstances. Having discovered the content of a person's constructs from what he says and does, the clinician explains his behavior by showing that the way he in fact acts in various situations is just what would seem the right way to act if one followed the rules he follows. If the person says or does something that cannot be explained this way—i.e., the constructs he is taken to be using do not lead to a calculation that would make this seem right under such circumstances —the explanation has been falsified by the facts.

Since people in their words and deeds not only deceive others, but often deceive themselves about the reasons for their choices, it may be very difficult to discover just how a person really construes. To discover this is the *empirical* task of the clinician and the *Psychology of Personal Constructs* is addressed to that task. Kelly (1955) tells the clinician:

> If you do not know what is wrong with a person, ask him; he may tell you. The clinician who asks such a question will have to be prepared to do some careful listening [pp. 322–323].

Kelly's whole theory, the various "corollaries" as well as related techniques like the REP test and the analyses of self-characterizations, may be regarded as a *guide to such "careful listening"*; it shows the clinician *what to look for* in order to discover "the full import of his client's thinking [pp. 71–72]"—i.e., why he construes what he does as the right thing to do. The clinician who discovers this can help his client because "from the standpoint of the psychology of personal constructs, psychological disorders can be traced to characteristics of a person's construction system [p. 832]." Thus for Kelly a psychological disorder "is not explained simply as the inevitable outcome of any past which happens to be unfortunate. It relates to one's present stance as he faces the future [p. 835]." Anxiety, guilt, hostility, aggressiveness, and various personality disorders are interpreted by Kelly as a function of the way a person construes, the way he uses self-defeating and deceiving rules in trying to cope with life (Ch. 16 & 17). Psychological illness thus becomes something that differs in kind from physical illness—not so much something that "happens" to a person as a causal consequence of "external forces bearing upon him," as something he himself does by the way he construes. And psychotherapy becomes helping a person to "reconstrue."

EXPLANATION VERSUS PREDICTION

But what about predictions? Kelly (1955) rightly points out that when the clinician understands the way people construe he has "grounds upon which [he] can make predictions regarding how people will act [p. 67]." If we understand the way someone construes then we can predict, e.g., that he is likely to classify another person as being of Type X and so to anticipate X-ish behavior from him; and that, since he construes Y as the right thing to do about X-ish behavior, he is likely to relate himself to this person by doing Y. So if someone follows certain rules then we can expect certain regularities in his behavior and thus make predictions. Yet the predictions which we base on the *rules* a person is taken to follow in deciding what to do differ from those based on nomothetic laws.

To see the difference, note that when laws are used to explain physical phenomena the criteria for deciding whether or not various phenomena are "similar," and thus capable of subsumption under the same nomothetic law, are those of the *observer*—what makes falling apples and the motions of the planets similar phenomena when we are doing dynamics is Newton's theory of motion. Since the observer's theory specifies the phenomena to which it is applicable, predictive failure shows that the theory on which the prediction was based *must* be faulty. This is also true when we are dealing with animal behavior because *we* set up the defining conditions for, e.g., a certain S or R class (Meehl, 1954, p. 42). But the situation changes when we explain human behavior in terms of the rules a person follows in deciding what to do. For then the criteria of similarity are not those of the observer but those of the *agent;* it is not what *we* regard as, e.g., "dominating" behavior, but what *he* regards as such behavior that determines how he will act. Since understanding how he construes is understanding the rules he follows in classifying people and deciding what to do about them, predictive failure will, of course, make us wonder whether he really construes the way we thought. But it will not compel us to withdraw the claim that this is how he construes. For it may be that, though he usually construes the way we think (i.e., classifies people on the basis of these rules, etc.), he had some reason—to be explicated in terms of the rest of his personal constructs—for construing differently in the situation where our prediction failed. Because *he* is the one that must apply the rule that determines "similarity," it is always possible that he had some reason for not seeing the situation as similar to others in the way we predicted he would. Since the prediction was not based on a nomothetic law whose "end terms" can be defined by the observer's theory without reference to the way that person himself construes, predictive failure will not show that there *must* be something wrong with the "theory" on which the prediction was based.

Consider an example: Meehl (1954) argues that the only way we can "prove" that a certain "way of talking" on the part of a patient is associated with hostility towards his sister is by carrying out a "statistical analysis on the verbal protocols of the individual case [p. 81]." But what constitutes a similar way of talking when exactly the same words said in a different context, or tone of voice, may be a different way of talking? The clinician will have to specify criteria for identifying similar ways of talking before he can carry out his statistical analysis. Failure to find a significant correlation between what *he* has specified as similar protocols and hostility may mean either that these protocols are not associated with hostility, or that from the patient's point of view they are not all similar. For it is not the clinician's criterion but the patient's own construal (i.e., *his* use of rules) which determines whether or not what he says on one occasion is really similar to what he says on others. Since statistics cannot settle this issue they cannot be the final court of appeal that settles claims about how a person really construes.

This leads to my final point: The statistical (predictive) approach and the personal construct approach have such different aims that they cannot be significantly compared. If we are interested in *predicting* behavior, then statistics may be a more reliable guide than personal constructs; but if we are interested in *explaining* human actions then personal constructs cannot be replaced by statistics. For predicting and explaining are different, and success in the former is not a criterion of success in the latter. Careful analysis of the way the tides behaved in the past enabled the ancient Babylonians to predict their future occurrences with accuracy; but they could not explain the occurrence of the tides. Newton's theory of gravitation provided an explanation, but it did not lead to better predictions. In fact, most tidal predictions are still based on statistical analyses of their past behavior rather than on Newton's theory. Though "positivistic" thinking about science was dominated by the idea that there is no essential difference between explanation and prediction, more recent studies of the logic and history of science have brought this idea under sharp attack even with respect to the physical sciences (Hanson, 1958, 1959; Toulmin, 1961). Surely this may encourage the clinician to point out that, since his primary aim is to understand his patient with a view to helping him, he is doing something different from the actuarian, something that cannot be relevantly evaluated in the same terms.

For suppose a clever statistician, who knows nothing of chess, were to observe the behavior of chess players. He would soon notice certain regularities in the way pawns, etc., move; he would also notice regularities in the way a particular player moves one figure rather than another (e.g., when besides the king he has only a rook and a bishop, under what conditions does he move the former rather than the latter?). Given time and a good calculating machine, he could learn to predict from the configuration of pieces on the board what "move" this player is likely to make. But predictive success would not show that he *understands* the move—indeed, he does not even understand that it *is* a move in the game of chess! It is simply irrelevant to compare what this statistician is doing with what is done by someone who wants to *make sense* of the chess player's behavior by showing, in terms of the rules of chess and the strategy followed by this player (e.g., the kind of end-game he was trying), why he construed this as the right move to make. If we can explain the moves made by a chess player in terms of the rules and strategic principles he follows, we will also be able to predict what move he is likely to make in certain circumstances. Still, what we are doing differs from what is done by the statistician interested in predicting behavior.

Philosophers influenced by Wittgenstein's (1953) discussion of rules have recently argued that since the social and historical sciences deal with rule-following human behavior they require a model of explanation which differs from that of the physical sciences (Dray, 1957; Winch, 1958). I have tried to develop a similar argument for clinical explanations based on personal constructs. Nothing I have said is intended to detract from the importance of nomothetic approaches to the study of behavior. But my argument is a defense of an idiographic approach to clinical activity in that it has tried to show that the use of personal constructs to explain individual behavior is different from, and independent of, the use of general laws.

REFERENCES

Dray, W. *Laws and explanations in history.* London: Oxford Univer. Press, 1957.

Falk, J. L. Issues distinguishing idiographic from nomothetic approaches to personality theory. *Psychol. Rev.,* 1956, **63**, 53–62.

Hanson, N. R. *Patterns of discovery.* Cambridge: Cambridge Univer. Press, 1958.

Hanson, N. R. On the symmetry between explanation and prediction. *Phil. Rev.,* 1959, **68**, 349–358.

Kelly, G. A. *The psychology of personal constructs,* Vol. 1, 2. New York: Norton, 1955.

Kelly, G. A. Man's construction of his alternatives. In G. Lindsey (Ed.), *Assessment of human motives.* New York: Grove Press, 1960. Pp. 33–64.

Kelly, G. A. Europe's matrix of decision. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation: 1962.* Lincoln: Univer. Nebraska Press, 1962.

McClelland, D. C. The psychology of mental content reconsidered. *Psychol. Rev.,* 1955, **62**, 297–314.

Meehl, P. E. *Clinical versus statistical prediction.* Minneapolis: Univer. Minnesota Press, 1954. Ch. 6.

Mischel, T. Psychology and explanations of human behavior. *Phil. Phenomenol. Res.,* 1963, **23**, 578–594.

Poch, S. M. A study of changes in personal constructs as related to interpersonal prediction and its outcomes. Unpublished doctoral dissertation, Ohio State University, 1952.

Toulmin, S. *Foresight and understanding.* Bloomington: Indiana Univer. Press, 1961.

Winch, P. *The idea of a social science.* London: Routledge & Kegan Paul, 1958.

Wittgenstein, L. *Philosophical investigations.* Oxford: Basil Blackwell, 1953.

(Received July 6, 1962)

Psychological Review
1964, Vol. 71, No. 3, 193-218

THE ORIGIN OF SELF-CRITICISM[1]

JUSTIN ARONFREED
University of Pennsylvania

The learning of self-criticism is represented as the prototype of a form of internalization in which the child replicates certain components of the punishments to which it has been previously exposed. The stimulus properties of these components, first presented in the responses of a model, attain reinforcement value when they come to operate as signals for attenuation of the child's anticipatory anxiety. The effects of a variety of experimental socialization paradigms are combined with an analysis of naturalistic socialization to illustrate the mechanisms in the acquisition of self-criticism. Some extinction paradigms are also used to examine the sources of its persistence. Both the motivating and the signalling functions of social punishment are described as being essential to its reproduction, while the rewarding characteristics of the model are seen to bear only a very tangential relationship to the phenomenon. The motivational antecedents and consequences of self-criticism are considered to be independent of the cognitive substance that defines its evaluative character.

The reproduction of the punitive reactions of socializing agents is a readily observable phenomenon in the behavior of young children. Children often show punitive responses to their own actions, and they sometimes recreate their parents' disciplinary actions in exact motor detail. As language and cognition increasingly take the place of action, social punishment becomes most extensively and persistently reproduced in the form of verbal and evaluative responses. It is very common, for example, to see a young child react to its misdemeanors by saying aloud to itself: "You're a bad boy!" or "That was bad!" Such responses become more covert as the child's symbolic equipment expands. But people do continue beyond childhood to frequently direct toward themselves (and toward others) various components of the social punishment to which they have been previously exposed. These self-punitive reactions belong to a class of behavior whose social learning represents one important type of imitation or identification. They are also of considerable interest because they seem rather paradoxical when approached in terms of our conventional concepts of motivation and reinforcement.

My purpose in this paper is to examine the origin of self-criticism as an illustrative case of the learning and motivational processes which more broadly underlie the adoption of a model's punitive responses. The conditions of aversive learning under which children acquire the use of the label "blue" to refer to their own actions, when the label has been repeatedly used as a component of punishment by a socializing agent, will be demonstrated with a series of experimental socialization paradigms. Evidence bearing upon the resistance to extinction of the self-critical response will also be presented. The results of the paradigms will be used, together with an analysis of naturalistic socialization, to advance the view that self-criticism is acquired because it actually

---

[1] The author is indebted to two of his colleagues, F. W. Irwin and R. L. Solomon, for their kind readings of an earlier version of this paper. He also wishes to thank Selma Roseman for her capable research assistance in collecting the experimental data.

reduces anxiety, and that it does so precisely because evaluative labels were originally subsumed under or contiguous with external punishment. The acquisition of self-punitive responses in general will be described as having only a very limited and indirect relationship to the rewarding characteristics of a model. Finally, it will be argued that self-criticism reveals the essential motivational antecedents and consequences of any internalized evaluative response to a socially defined transgression, regardless of the nature or extent of the cognitive substance of the response.

The evaluative judgment of one's own actions has been traditionally regarded as the foundation of internalized conduct (Freud, 1936, Ch. 8; Hartshorne & May, 1928; Piaget, 1948; Sears, Maccoby, & Levin, 1957, Ch. 10). Many self-evaluative responses must be derived, of course, from social learning based on reward. But the tendency of children to independently replicate the punitive and prohibitive functions of socialization makes a distinct contribution, and one that is at least comparable with the effects of social reward, to the distant relationship that human behavior often bears to what would seem like more inherently preferred responses. This replicative effect of social punishment is clearly displayed in the internalized dispositions which lead people to impose upon themselves both restraint and unpleasant consequences of their own actions. In many contemporary psychological conceptions of internalization (Miller et al., 1960, Ch. 5; Sears et al., 1957, Ch. 10; Whiting & Child, 1953, Ch. 11), self-criticism is viewed as the mediator of other reactions to transgression, and also of control over the occurrence of transgression, when such responses can be elicited in the absence of direct external surveillance. It is, in fact, commonly assumed that the internalized consequences of social punishment require the operation of explicit standards or values which have been adopted from those of a model.

There is actually considerable evidence to question the assumption that the adoption of a model's values or standards is requisite to internalization. The casual observation of the first appearance of internalization in preverbal children makes it particularly obvious that internalized reactions to socially punished acts do not necessarily imply the presence of self-critical cognition. Even when they do not yet have the equipment for social evaluation, young children frequently show signs of anxiety and anxiety-reducing maneuvers following a transgression that has not been apparently open to external observation. Some recent work with dogs (reported by R. L. Solomon in Mowrer, 1960b, pp. 399–404) also suggests that internalization does not require a highly developed cognitive base. Moreover, numerous studies have indicated that internalized responses to transgression show great variability, among both children and adults, in the nature and extent of the cognitive mediating processes on which they rest (Allinsmith, 1960; Aronfreed, 1960; Kohlberg, 1963; Whiting, 1959). Even self-corrective responses, such as confession or reparation, have been found to occur frequently without evaluative standards being applied to a transgression (Aronfreed, 1961). Nevertheless, self-criticism surely deserves special attention among the internalized consequences of social punishment. In addition to its significance for the phenomenon of identification, it also provides verbal and cognitive signals through which other social responses can be mediated with a persistence and generalization that would be impossible if they were forced to depend only on the immediate stimuli of the transgression itself and of its external context.

ACQUISITION

Actions (and, sometimes, intentions or other intrinsic precursors of action) may be defined as transgressions to the extent that they have been followed, in the child's experience, by any form of social punishment. Social punishment includes, of course, not only the application of concrete noxious stimuli, as in physical punishment, but also withdrawal of affection and a wide variety of other kinds of aversive events. When an action is defined as a transgression, through its aversive social consequences for the child, its response-produced cues become signals for a conditioned drive state that can be given the general designation of anxiety. The anxiety may have, however, qualitative variations which reflect its cognitive context and possibly the nature of the original punishment. It is the invariant motivational source of the child's suppression of transgression, and also of a number of different responses to the occurrence of transgression, all of which share in common their instrumental value for reducing the anxiety. The instrumental status of these responses is derived from the attachment of anxiety attenuation to their discriminant stimulus properties, as a result of certain significant reinforcement contingencies which are present in the socialization process. The responses may be said to be internalized to the extent that the anxiety which motivates them has come to be directly elicited by transgression and independent of either external punishment or social cues which portend punishment.[2] Such instrumental responses are quite variable with respect to the precise patterns of learning through which they become self-reinforcing for the child. Recent studies in the experimental induction of self-criticism and reparation suggest, for example, that these two reactions to transgression are different in the specific contingencies required for their acquisition and in their dependence on evaluative cognition (Aronfreed, 1963; Aronfreed, Cutick, & Fagen, 1963). They cannot be regarded as alternative expressions of a unitary evaluative and punitive agency such as "conscience" or "superego."

The anxiety-reducing value of some internalized responses to transgression, such as confession or reparation, may be acquired because the responses themselves originally received direct external reinforcement through the prevention or arrest of punishment (or through the removal of social cues which portended punishment). But other responses can produce cues to which anxiety reduction has become attached even when the responses have not occurred during the initial socialization process. The most prevalent instances of such responses are those in which the child reproduces components of its previous punishments. Under the ordinary circumstances of child rearing, many components of relatively brief social punishments are sufficiently close to the anxiety termination that follows them to take on substantial power as cues for the reduction of the anticipatory anxiety that has already become directly associated with transgression. The anxiety-reducing function of these components of punishment is central to an understanding of self-criticism. Quite aside from its cognitive requirements, self-criticism is a replication of the verbal medium of punishment. When a child's punishment incorporates verbal criticism—the application of verbal labels or standards to its behavior—the criticism, like any component of punishment, may attain the significance of a cue signaling the attenuation of the antecedent anxiety that occurs in the interval between transgression and punishment. Following subsequent transgressions, the child itself can make the critical response and thus provide itself with anxiety-reducing cues. The cue aspects of the self-critical response, initially presented in the punitive behavior of a model, thus acquire reinforcement value through their conjunction with anxiety reduction, without the response having to be first overtly emitted by the child and then externally reinforced. A component of punishment becomes preferred to the experience of anticipatory anxiety.[3]

---

[2] In defining internalization, one might make a more subtle distinction between whether the anxiety occurs merely in the absence of punishment or whether it occurs in the absence of other external aspects of the original punishment situation (such as perceived observation by others). While the latter condition is a more stringent criterion of internalization, there are significant advantages in providing a broader base of definition. It is very difficult to know whether a person's *perceived* environment includes the possibility of external observation. And once reactions to transgression become independent of punishment, they may take much the same form regardless of the actual presence or absence of observers.

This description of the learning of self-criticism specifies that it is established as the consequence rather than the source of the negative affect that we experience following a transgression. The anxiety-reducing function of reproduced punishment is sometimes compellingly visible in young children who, even though they are unaware of being observed, may show evidence of acute discomfort when they misbehave and then apparent relief after they have criticized or punished themselves. The introspective perception that the self-critical response sometimes precedes unpleasant feelings may be phenomenologically misleading if it is used to judge the validity of the mechanism of acquisition set forth here. The response, once firmly acquired on an instrumental basis, may begin to occur directly to the cues of transgression with a latency shorter than that with which anxiety reaches substantial intensity. Its maintenance would continue to be dependent on anxiety reduction, however, since the response would tend to be repeated once the anxiety had attained sufficient strength. Recent studies (Brady, Schreiner, Geller, & Kling, 1954; Brush, Brush, & Solomon, 1955) emphasize the importance of distinguishing the position of anxiety in the acquisition of an aversive response from its position in the maintenance of the response, and Solomon and Wynne (1954) have analyzed this problem specifically in terms of latencies.[4]

[3] The anxiety-reducing cue value of punishment may also be useful in understanding why some individuals rely primarily on punitive external events to define the consequences of transgression (Aronfreed, 1961; Kohlberg, 1963). Such individuals may be those who have been given sparse verbal and cognitive equipment or who have not found it reinforcing to use their own behavioral resources in response to transgression (for example, young children or members of lower status groups).

[4] The child's repeated or sustained emission of self-criticism following a transgression may be attributable to the persistence of the anxiety response. Its extension in time is very probably a function of the marked variability in the duration of punishment used by socializing agents and in the temporal position of their critical labels with respect to onset and termination of both transgression and punishment. It should also be noted that verbal criticism is often used as a warning signal that punishment will follow if an incipient transgression is carried out or if an ongoing transgression is not corrected. Consequently, the self-critical response, though motivated by ante-

The learning of self-criticism embraces, then, two processes: (*a*) the attachment of anxiety to the response-produced cues of transgression through their repeated association with punishment; (*b*) an accrual of the intrinsic reinforcement value of the self-critical response through the inclusion of its cue properties within external punishment in such a way as to make them contiguous with the attenuation of anxiety. The use of two distinct processes to account for the establishment of aversive responses is not uncommon (Mowrer, 1960a, Ch. 3–7; Solomon & Brush, 1956). But the second process described here has certain features which are not ordinarily found in the experimental literature on aversive learning. The learning of an aversive response is typically effected by having the response first occur and then receive direct external reinforcement through avoidance of, or escape from, punishment. In contrast, self-criticism is representative of a class of responses whose acquisition of functional value does not require either that they occur overtly or that they have discriminable effects upon external reinforcing events. Instead, their value is accumulated as their cue properties are repeatedly presented independently, through the external social behavior of a model, in a temporal position that permits them to serve as signals for the reduction of an anticipatory state of anxiety. The child's subsequent emission of these responses therefore becomes immediately self-reinforcing upon their first occurrence. Support for such a conception is found in demonstrations that neutral stimuli may acquire either motivating or reinforcing properties, according to their temporal association with aversive states, while relevant instrumental responses are not being made by the organism (Gleitman, 1955; Montgomery & Galton, 1955; Solomon & Turner, 1962).

There are some convincing indications that direct instrumental learning can bring an organism to make responses which correspond to those of a model (Church, 1957; Darby & Riopelle, 1959; Miller & Dollard, 1941). And Hill (1960) has recently specifically attempted to describe the learning of self-criticism by using punishment avoidance or escape as the external reinforcement of an initially overt response. But one would expect behavior that is established through direct social reinforcement, even when it happens to be similar to the behavior of a model, to have the kind of latitude usually permitted to responses whose instrumental consequences are originally externally defined. Social reinforcement is not ordinarily so selective as to be contingent on an extremely restricted form of response. Self-criticism, however, has the earmarks of the truly imitative response. It reproduces fairly precisely the stimulus properties of a model, and thus allows a strong presumption that its reinforcement is rather narrowly bound to those properties.

It also seems implausible for other reasons that direct external reinforcement makes any substantial contribution to the initial emergence of self-criticism. Of course, the overt verbalization of the response by young children, once it has been established, might result in the withholding of punishment or even in reward by their parents. But socializing agents can hardly consistently reinforce a response that is generally not apparent in behavior (indeed, they must often be administering punishment following self-criticism, whether it be implicit or

---

cedent anxiety, may not be sufficiently anxiety-reducing unless it is accompanied by inhibition or corrections of deviance.

overt). In view of the very limited opportunity for such a contingency that one sees in naturalistic child rearing, it would certainly seem a cumbersome and lengthy process to externally reinforce a tendency in the child to reconstruct a component of its previous punishment in order to avoid the delivery of punishment that is externally imminent.

Anyone who has observed the vigor and suddenness with which self-criticism characteristically becomes observable in young children cannot escape the inference that the response already has considerable strength at its first verbalization. This prepotency makes it necessary to go beyond the assertion that the stimulus properties of the response attain anxiety-reducing value without its overt occurrence. We must also assume that the critical label is implicitly emitted by the child during the original punishment learning. The response would then, before it became overt, have already been strengthened through the reinforcement value gradually attached to its intrinsic cue aspects. Given the cognitive capacity of children, the presence of the label in their response repertoire, and the possibly close connection between the receptive and productive functions of language, we might well expect their covert repetition of any label used by a socializing agent.

Finally, it should be noted that this conception of the learning of self-criticism touches upon an issue of more general significance. There are currently a number of inconsistent experimental findings on the question of whether stimuli can become positive reinforcers through their association with the termination of aversive states. The phenomenon has apparently been observed in some animal studies, under the appropriate conditions (Buchanan, 1958; Goodson & Brownstein, 1955; Montgomery & Galton, 1955), but other studies indicate primarily the difficulty of its demonstration in the laboratory (Littman & Wade, 1955; Mowrer & Aiken, 1954; Nefzger, 1957). Some of these experiments, however, do not really duplicate the conditions of social learning under which we could expect to observe such an effect. As Beck (1961) has recently pointed out, attempts to test the reinforcing properties of stimuli which have been contiguous with termination of anxiety or punishment may not be successful if they are carried out in situations in which neither anxiety nor pain is present. The acquisition of self-punitive responses does depend, of course, on the presence of anxiety during both socialization and subsequent occasions when the child itself can reproduce the components of punishment.

Naturalistic and Experimental Socialization

The view that self-criticism is acquired because it attenuates an aversive state does not require that the external criticism of socializing agents occur at a particular point in the course of punishment. What is crucial to the child's reproduction of punishment is the temporal position of its components with respect to onset and termination of anxiety, and not their position with respect to onset and termination of the punishment itself. The usual duration and intensity of social punishment is such that even a component at its onset might acquire some anxiety-reducing significance, provided only that there is sufficient time after the transgression for the arousal of anxiety. When children are quite young and under close parental control, punishment often follows fairly quickly upon transgression. But it does not follow so quickly as to prevent an interval of anticipatory anxiety, particularly since the anxiety comes to be elicited by the incipient precursors of transgression and not merely by its commitment. And, as children become older and more mobile, the interval between transgression and punishment tends to lengthen. Consequently, despite the great variability of parents in their timing and ordering of components of punishment, any component could assume some anxiety-reducing value.

To the extent, then, that parental punishment induces internalized anxiety and is administered in a verbal and cognitive medium, some acquisition of self-criticism by the child would seem inevitable. A close examination of the learning of self-criticism therefore requires artificial conditions under which the verbal components of social punishment are varied in their anxiety-reducing significance. The critical labels would presumably not become anxiety-reducing, for example, if they were to follow immediately upon the intrinsic cues of transgression (so that anticipatory anxiety would be negligible) and were consistently the first element in a punitive reaction of some duration. Conversely, one would expect the labels to have maximal anxiety-reducing value if they were always to occur at the termination of punishment. The acquisition of self-criticism under experimental conditions might also be facilitated if the onset of punishment (but not the critical labels) were close to the completion of transgression. Such a temporal contingency probably occurs quite often in the early phases of socialization, when the young child's behavioral limitations make it subject to fairly immediate punishment following forbidden behavior, and it may make a potent contribution to the anxiety that motivates self-punitive responses. Animal studies (Davitz, Mason, Mowrer, & Viek, 1957; Kamin, 1959) have indicated that short intervals between response-produced cues and the occurrence of punishment facilitate the acquisition of anxiety-reducing responses. These temporal variations do not occur with any consistency, of course, in the behavior of parents. But they can be brought under control in experimental situations.

An experimental induction of self-criticism ought further to be enhanced if the model's critical labels coincide with the termination of the response-produced cues associated with the transgression. In a naturalistic setting, parental punishment and criticism often interfere with an ongoing transgression, and thus eliminate its intrinsic cues and the anticipatory anxiety already attached to them. The child's reproduction of cue components of punishment might well be supported by this effect, since responses have been shown to have an advantage in learning when their stimulus properties coincide with the termination of a conditioned aversive stimulus (Brush et al., 1955; Church, Brush, & Solomon, 1956; Mowrer & Lamoreaux, 1942). A rapid acquisition of value by the critical labels of a socializing agent, when they coincide with termination of transgression, might also reflect the absence of any delay of the reinforcement represented in anxiety reduction. When the anxiety-reducing significance of a component of punishment becomes discriminable to a child, there is no delay between the parent's use of that component and the occurrence of anxiety reduction. Neutral stimuli have been shown to attain more powerful secondary reinforcing properties when there is a short interval between their presentation and the original reinforcement (Jenkins, 1950; Schoenfeld, Antonitis, & Bersh, 1950).

There is one other feature of naturalistic socialization whose effect upon the internalized consequences of punishment is difficult to assess, because it is relatively restricted in variation and inseparable from the total parent-child relationship. It is the parental nurturance and affection that serves as the background for punishment learning. Many psychologists have felt that the gratifications and rewards provided by parents are an essential ingredient of the child's adoption of their behavior and values in any form. Freud (1936, Ch. 8), for example, regarded the motivation to reproduce a loving and rewarding parental model as one of the major sources of identification. And other theorists (Bronfenbrenner, 1960; Miller et al., 1960, Ch. 5; Sears et al., 1957, Ch. 10; Whiting & Child, 1953, Ch. 11) have attempted to extend this motivation to the internalization of social punishment.

The most detailed rationale for the generalized effects of nurturance upon internalization is found in the work of Sears et al. (1957, Ch. 10) and of Whiting and Child (1953, Ch. 11). The child is presumed to be broadly motivated to identify with its parents in order to reproduce the pleasurable experience associated with the stimulus aspects of their behavior. In making responses like those of the parents, particularly when they are absent or have withdrawn affection, the child provides itself with stimuli which have become powerful secondary reinforcers through their conjunction with parental rewards and gratifications. Thus, the child's application of the parents' critical evaluative responses to its own actions represents, according to this view, an instance of a more general tendency to reproduce many kinds of parental behavior. And when the child critically evaluates its own behavior, it experiences unpleasant feelings, just as when it was previously exposed to external criticism; that is, self-criticism is regarded as the original internalized source, rather than the consequence, of the anxiety that follows a transgression.

The use of experimental paradigms affords an opportunity to compare the contribution of a socializing agent's nurturance to the induction of self-criticism with that of the anxiety-reducing value of cue components of punishment. If the internalization of social punishment were a function of the model's nurturance, or of the withdrawal of affection conveyed in the model's punishment, one would expect that the child would be likely to adopt the criticism of an otherwise nurturant model in a punishment learning situation. Its adoption of the model's criticism would presumably reflect a more general motivation to reproduce the stimulus aspects of the model, and should accordingly not be contingent on the timing of the criticism with respect to onset and termination of anxiety.

Four Socialization Paradigms

In order to examine the origin of self-criticism at close range and under controlled conditions, four experimental paradigms were devised, in each of which the experimenter repeatedly used the label "blue" to refer to a specific response whenever it was made by the child in the course of 10 training trials. The label "red" was used to refer to alternative responses. A contrived task made it possible to control the frequency of the blue response and to use an artificial external signal to indicate its occurrence. In three of the paradigms, the blue response was defined as a transgression by punishment that consisted of the experimenter's sharp disapproval and removal of a candy that had been given to the child. The punishment and the external signal were given a common onset and termination. In the fourth paradigm, the blue response was not associated with punishment. None of the training paradigms offered the child an opportunity to overtly verbalize the evaluative labels used by the experimenter. However, all of the paradigms were uniformly followed by three test trials designed to observe whether the child would use the labels to refer to its own actions when the socializing agent's role was temporarily eliminated.

The external signal and the punitive reaction of the experimenter were initiated, during the three punishment paradigms, immediately upon occurrence of the blue response. Since the punishment had a number of components and some extension in time, the label "blue" could be placed in different positions so as to produce extreme variations in its timing with respect to onset and termination of both the punishment and the external signal of the forbidden response. In the first punishment paradigm, the experimenter placed the label at termination of signal and punishment, the position at which it was expected to assume maximal anxiety-reducing value. In a second paradigm, the label was used at the very onset of signal and punishment, where there was virtually no time for anticipatory anxiety to develop. The experimenter's general (nonpunitive) demeanor, in both of these first two paradigms, was neutral and somewhat distant.

In the third punishment paradigm, the experimenter displayed considerable warmth and approval toward the child, except when administering punishment, and expressed both physical and verbal affection. This procedure was designed to create a nurturant context for socialization and an experience of nurturance withdrawal when the child was punished. The label "blue" was placed at the onset of punishment and signal, so that the effects of the nurturance itself could be compared with the effects, in the first paradigm, of maximizing the anxiety-reducing value of the label. The fourth paradigm was devised as a control to support the interpretation of the role of punishment in establishing the anxiety-reducing function of the label. No punishment was used, but the label occurred at the same point, the termination of the external response-signal, as it did in the first paradigm.

The basic design of these paradigms, and some of their other operational features, were arrived at on the basis of three central tactical considerations which emerged from preliminary studies:

1. Parents ordinarily get a number of "trials" in which to induce an internalized reaction to transgression, because young children commit the transgression on many occasions before internal controls are effective. But if the already socialized children who must be used in an experiment had their own intrinsic response cues to identify a punished act, they might inhibit the act well before a new self-critical response had been learned. Furthermore, their incipient response cues would make it impossible to have an experimental condition in which anticipatory anxiety was negligible. Accordingly, the blue response was designed to be intrinsically indistinguishable from other available responses. Its occurrence was signaled by a fairly loud buzzer, whose onset was immediately coincident with the response. This signal was, then, an externalized response-produced cue.

2. The punishment paradigms would be poor replicates of naturalistic socialization, and might arouse irrelevant competing responses, if the child were punished for transgression on every trial. Children do not always make the socially disapproved response.

Blue responses were therefore interspersed with red responses which were not punished and were actually rewarded in the nurturance paradigm. This procedure made the label "blue" a more discriminative stimulus, and also permitted a comparison of the child's adoption of two different kinds of labels.

3. Since self-criticism had to be observable on the test trials, a stimulus was required to elicit it in a verbal form. Moreover, the experimenter had to remain in the situation, in order to observe the response. At the same time, the experimenter's punitive presence had to be somehow perceived as temporarily absent or deferred, if the child were to have occasion and motivation to reproduce the experimenter's evaluative responses. These requirements for the test trials were met by having the experimenter: (*a*) appear to be absorbed in other matters and temporarily discontinue the socializing role, and (*b*) casually provide brief nondirective verbal-eliciting stimuli, so as to make the situation appropriate for a wide variety of verbal responses from the child.

Subjects

The subjects for the acquisition phase of the study were 89 girls drawn from the fourth- and fifth-grade populations of two public schools.[5] The number of subjects assigned to each of the four socialization paradigms was as follows: Labeling at Termination of Punishment: 27; Labeling at Onset of Punishment: 24; Nurturance (with labeling at onset of punishment): 26; Control: 12.

Procedure

Each child was taken by the experimenter (*E*), who was a female, from her classroom to the experimental room. The child sat in front of a small table, upon which was a piece of apparatus referred to as "the machine" and also a small box containing 15 Tootsie Rolls. The machine was simply an enclosed, black, metal cabinet, roughly 15 inches wide, 10 inches high, and 8 inches deep. In addition to a buzzer located within the machine, the following items were embedded in its inclined face: an ac voltmeter, an electrical counter (with display), four small lights (two red and two green) ordered vertically along the left edge, and four locking lever switches running horizontally along the bottom edge and spaced about 3 inches apart. The primary purpose of these units was to convey to the child that the machine was a formidable device. Under each unit was a decal purporting to refer to a complex instrument within the cabinet, such as "ACO Narrow Band Monitor," etc.

In order to generate the operation of the machine on each trial, the child was given the task of pushing down the number of levers that indicated her guess as to how many of four hidden dolls were looking in her direction. The nature of the blue response was vaguely defined for the child as having to do with the way in which the levers were pushed. However, the actual occurrence of transgression was controlled by *E* through a hidden silent mercury switch that monitored the activation of the units in the machine. The children tended to push first the lever on their far left. This lever was therefore made inactive, so that the blue response would not always occur at the beginning of a guessing sequence. The other three levers were equivalent in that any one of them could activate the buzzer and the other units. When activated, the indicator on the voltmeter would swing, the counter would click and turn over once, and one, two, or three of the four lights would go on (according to which of the levers had been pushed). The child was told that there would always be at least two dolls looking toward her, in order to guarantee the occurrence of the blue response on appropriate trials.

Approximately 6 feet behind the child, *E* sat at a second table, upon which was a high black cardboard screen. Behind the screen were the four identical small dolls and the mercury switch. After *E* and *S* were seated, *E* used one of four paradigms, each of which consisted of initial instructions and a sequence of training trials designed to make it difficult for *S* to anticipate events on any given trial. Blue responses occurred on Trials 2, 5, 6, 8, and 9, and red responses on Trials 1, 3, 4, 7, and 10. Each trial began at roughly 8 to 10 seconds after the previous one ended.

---

[5] The author is grateful to a number of administrators and teachers in the Philadelphia school system, whose cooperation made this study possible.

Labeling at Termination of Punishment (LTP). The *E*'s initial instructions were as follows:

> Here's what we're going to do today. Back here, I have four dolls (*E* held up dolls, then replaced them behind screen). I'm going to place some of them looking at you from behind here and some looking away (*E* appeared to handle dolls behind screen). You're supposed to guess how many are looking at you. There are always at least two looking at you, but sometimes there might be more. Sometimes three of them might be looking at you, and sometimes all of them. Each time, you guess, and then I change them, and then you guess again.

The *E* then walked over to *S*'s table and said:

> Now what I really want to see is how you work with a machine. The way you guess is: you tell the machine what you think is the answer. The machine is automatic. When I set the dolls, the machine is set, too, with the right answer. Here's the way the machine knows whether you've guessed how many dolls are looking at you: you push down as many levers as you think there are dolls. If you think there are two dolls looking at you, push two levers (*E* pushed down first two levers from the left. The mercury switch was open at this point, so that none of the levers were activated). If you think there are three dolls looking at you, push three levers (*E* pushed down third lever). And if you think there as four, push four (*E* pushed down last lever). Remember that you always push at least two, and leave them down. Don't push them back up. The machine doesn't show you whether you guessed right, but it tells my instruments, so I know. And when we're all done, I'll tell you how many times you guessed right. Do you understand?

After *S*'s indication of understanding, *E* reset all levers, returned to her own table, and, while appearing to handle the dolls, closed the mercury switch. Then *E* returned to *S*'s table and continued:

> Now the most important thing is the way you use the machine, the way you push down the levers. I've done this with a lot of girls, and I've found out that there are two different kinds of girls who use the machine in two different ways. I don't know what to call them, so I just call one kind the girls who do it the *red* way and the others the ones who do it the *blue* way.
>
> Over here is a pile of Tootsie Rolls for you (*E* removed Tootsie Rolls from box and placed them in front of *S*). When you use the machine the *red* way, it just tells my instruments over there what you've guessed (*E* pointed to her own table). You push the lever down with your finger from the top (*E* demonstrated, carefully pushing down the inactive lever). Now the machine has a buzzer that sometimes goes on (here *E* ostensibly adjusted something at the back of the machine)—it sounds like this (*E* pushed down the second lever, activating the buzzer and the other units in the cabinet). You can hear the buzzer and you can see what's happening (*E* indicated voltmeter and lights). When that happens, you lose some of your Tootsie Rolls. I take them away and put them in this empty box here, because you did it the *blue* way (as *E* said the word "blue," she reset the levers, thus terminating the buzzer and the activation of the other units).

The *E* then returned to her seat behind *S*, and the training trials began. To begin the first trial, *E* said: "All right, guess how many dolls are facing you. Just push down the number of levers you think is right." To begin subsequent trials, *E* simply said: "All right, you can guess now." On a nontransgression trial, the mercury switch was open. *E* merely rose from her seat and said: "Uh-huh, you did it the *red* way." She then walked over to *S*'s table, reset the levers, and returned to her own table to begin the next trial. On a transgression trial, the mercury switch was closed, and the buzzer and other units were activated. *E* rose immediately, as the buzzer went on, and said: "No!" (fairly sharply). After walking quickly over to *S*'s table and looking with concern at the machine, *E* added: "You lose some of your Tootsie Rolls." The *E* then emphatically removed two Tootsie Rolls from *S*'s pile, put them in the box from which they came, and finally added: "Because you did it the *blue* way." Just as *E* spoke the word "blue," she simultaneously reset all of the levers that had been pushed down, so that the label was contiguous with the termination of both punishment and buzzer. Then *E* returned to her own table and began the next trial.

Labeling at Onset of Punishment (LOP). The procedure for this paradigm was identical, in most respects, to that described for the LTP paradigm above, except that the label "blue" was used at onset of punishment during training trials and occupied a corresponding position in the demonstration that accompanied instructions. The modification of instructions began at the point at which the Tootsie Rolls were introduced:

> Over here is a pile of Tootsie Rolls for you. When you use the machine the *red* way, it just tells my instruments over there what you've guessed. You push the lever down with your finger from the top (*E* demonstrated on inactive lever). When you use it the *blue* way, the machine has a buzzer that goes on—it sounds like this (*E* pushed down the second, active lever). You can hear the buzzer and you can see what's happening. When that happens, you lose some of your Tootsie Rolls. I take them away and put them in this empty box here (then *E* reset the levers and terminated the buzzer, but without using the label "blue" at this point).

Nontransgression training trials were treated in the same way as in the LTP paradigm. On transgression trials, however, *E* rose and said, immediately at the onset of buzzer and punishment: "You did it the *blue* way—No!" Then *E* walked over to *S*'s table and added, while looking at the machine: "You lose some of your Tootsie Rolls." *E* removed two Tootsie Rolls from *S*'s pile, put them into the box, and terminated buzzer and punishment by resetting all levers. The label "blue" was not used, of course, at this termination point.

Nurturance (with labeling at onset of punishment). The procedure in this paradigm was basically the same as that used in the LOP paradigm. Here, however, *E*'s manner and voice conveyed considerable warmth and approval of *S* (except on transgression trials), in contrast to her generally neutral demeanor in all of the other paradigms. The modifications of procedure entailed intersticed expressions of verbal and physical affection during the instructions and the nontransgression trials. Thus, after the initial distinction between the two kinds of girls who used the machine the red way or the blue way, and just before introducing the Tootsie Rolls, *E* interjected the following comment (in a smiling, relaxed fashion):

> Most of the girls in my own special machine class (*E* had been originally introduced to all *S*s as a "special teacher") are girls who do it the *red* way. As a matter of fact, you look something like one of them. I think I'm going to like you as much as I like my own girls (and *E* gently patted *S* on the head or shoulder).

On transgression trials, *E*'s behavior was identical to that described above for the LOP paradigm. On nontransgression trials, *E* rose and said, with maximal warmth and enthusiasm: "Good!—you did it the *red* way." The *E* then walked over to *S*'s table, reset the levers, and smilingly patted *S* on the head or shoulder before returning to her own table.

Control (labeling without punishment). The procedure for the control paradigm was similar in most respects to the procedure used in the LTP paradigms. There was, however, no punishment. The last brief section of the instructions was modified to inform the child of the possibility of punishment, so that this paradigm would be equated to others in having the general threat of punitive sanctions. The informational role of the buzzer was also emphasized somewhat more than it had been in other paradigms, in order to eliminate any punitive connotation in the association between the buzzer and the label "blue." After the initial introduction to the task, the modified section of the instructions continued as follows:

> Over here is a pile of Tootsie Rolls for you. I might take a couple of them away sometimes, depending on how you use the machine. The machine and I need to know whether it's the *red* way or the *blue* way. So we have a buzzer in the machine. When you use the machine the *red* way, it just tells my instruments over there what you've guessed (*E* pushed down the inactive lever). When you use the machine the *blue* way, we can tell that, too, because the buzzer goes on—it sounds like this (*E* pushed down second lever). You can hear the buzzer and you can see what's happening—because you did it the *blue* way (as *E* said the word blue, she reset the levers, terminating the buzzer).

On trials in which the blue response occurred, *E* rose, walked over to *S*'s table and said: "Uh-huh, you did it the *blue* way." As *E* said the word "blue," she terminated the buzzer. The *E*'s reaction to red responses was the same as in the LTP and LOP paradigms.

Test trials. Trials 11, 12, and 13 were test trials uniformly employed immediately after all of the socialization paradigms. The

S's lever choices on Trials 11 and 12 were followed by the buzzer, in order to test S's use of the label "blue" in reference to her action. To test for S's use of the label "red," S's lever choices on Trial 13 were not followed by the buzzer. The S's verbal responses on all test trials were recorded verbatim.

The E remained seated behind S on Trial 11, and appeared oblivious to the buzzer, having intentionally preoccupied herself with her recording sheets. If the child made no comment within 10 seconds after buzzer onset, E casually asked: "What happened this time?" This stimulus was used to elicit a verbal response from S. If S applied the label "blue" to her own behavior in any way, either spontaneously or in response to the eliciting stimulus, E then terminated the test trial in the same way that she terminated trials in the preceding training series (including the labeling and punishment, where the latter was appropriate). If S made a verbal response that did not include the use of the label "blue," E asked: "What do you mean?" After this last question, E terminated the test trial regardless of S's response. The punishment used in training was continued, in closing the test trial, so that the label "blue," if used by the child, would not receive a direct external reinforcement through punishment avoidance, since it was also to be tested on the subsequent trial (12).

The procedure for Trial 12 was the same as that for Trial 11, but was preceded by a statement intended to facilitate spontaneous responses. Before Trial 12 began, E said: "My eyes are getting tired from looking at these instruments. After you've guessed this time, you can tell me what happened yourself." The same procedure was also used on Trial 13, but the introductory statement used on Trial 12 was not repeated. After testing for S's use of the label "red," E concluded Trial 13 in the way appropriate to red responses on the preceding training series.

Closing Procedure. A closing procedure was used to assure S of the adequacy of her performance, to relieve any anxiety about the loss of Tootsie Rolls, and to invoke S's cooperation in not discussing the experiment with other children.

Results and Discussion

Certain features of the children's behavior during the training paradigms were strikingly apparent. They uniformly exercised great care in pushing the levers and showed obvious signs of discomfort upon the occurrence of transgression. Their pleased smiles in the nurturance paradigm also affirmed the experimenter's success in conveying affection. The power acquired by the buzzer was clearly visible on Test Trials 11 and 12 (except in the case of the control paradigm). The children were restless and tense, and most of them looked back expectantly at the now preoccupied experimenter. Although some children in the LTP paradigm made spontaneous labeling responses, the children generally verbalized the labels only in response to the experimenter's eliciting stimuli. This was true even on Trial 12, despite the initial statement that the children themselves could report on what happened, and was very possibly attributable to the effective functioning of the labels at an implicit level.

Numerous responses were given to the experimenter's verbal stimuli. Many children simply offered reportorial statements such as: "the buzzer went on," "I pushed the levers," "the lights went on," "I don't know," "you'll have to get up again," "I lose some Tootsie Rolls," etc. But only the relevant labeling responses, applied by the children to their behavior or to themselves, were used as evidence of self-evaluation. These generally took the form of: "I did it the blue (red) way!" or, less commonly, "I was blue (red)!" Table 1 shows, for each of the test trials, the relative frequencies of the appropriate labeling responses among the four paradigms. Chi-square values for comparisons of the frequencies are presented in Table 2. The observations on Test Trial 12 add only an indication of the stability of the responses seen on Trial 11.

The paradigm that is singularly effective in the induction of self-criticism

TABLE 1

FREQUENCY OF LABELING AND NONLABELING RESPONSES ON THREE TEST TRIALS

Response	LTP[a] ($N = 27$)	LOP[b] ($N = 24$)	Nurturance (LOP)[b] ($N = 26$)	Control[c] ($N = 12$)
Trial 11				
Blue label used	19	5	5	3
Blue label not used	8	19	21	9
Trial 12				
Blue label used	19	6	6	3
Blue label not used	8	18	20	9
Trial 13				
Red label used	9	6	7	3
Red label not used	18	18	18	9

[a] LTP = Labeling at termination of punishment.
[b] LOP = Labeling at onset of punishment.
[c] Control = Labeling without punishment.

is the one in which the socializing model uses the critical label at the termination of both the punishment and the externalized response-contingent signal of transgression. This finding strongly confirms the view that the child's reproduction of a component of social punishment rests on its original external position with respect to the onset and attenuation of anxiety. The fact that the blue label clearly attains functional value only in the temporal setting of the LTP paradigm, despite its not having been overtly verbalized during training, also points to the acquisition of self-criticism as an instrumental response that does not require direct external reinforcement. The infrequent use of the label among children exposed to a model who used it at onset of punishment and transgression signal, including those socialized in a nurturant context, makes it apparent that the reinforcing properties of the label are not to be assigned to any generalized effect of punishment. There is also no support for the view that self-criticism represents a broader tendency to reproduce the behavior of a model who is nurturant but withdraws nurturance in order to deliver punishment. The results of the experiment indicate that self-criticism is acquired as a specific response through

TABLE 2

CHI-SQUARE VALUES FOR COMPARISONS AMONG FOUR EXPERIMENTAL SOCIALIZATION PARADIGMS OF THE FREQUENCY OF LABELING AND NONLABELING RESPONSES

Comparison	Trial 11 (Blue label)	Trial 12 (Blue label)	Trial 13 (Red label)
LTP versus LOP	10.61**	8.73**	0.12
LTP versus nurturance (LOP)	11.99**	10.07**	0.01
LTP versus control	5.23*	5.23*	0.02
LOP versus nurturance (LOP)	0.05	0.03	0.01
LOP versus control	0.02	0.16	0.16
Nurturance (LOP) versus control	0.00	0.10	0.04

Note.—Chi-square values for 2 × 2 contingency tables (employing correction for continuity) based on frequencies in Table 1.

* $p < .05$ (one-tailed test).

** $p < .01$ (one-tailed test).

aversive learning. Its high probability of occurrence, when its stimulus properties have been independently presented as the terminal component of punishment, cannot be understood without reference to the mediation of anxiety and anxiety reduction.

The control paradigm eliminates interpretation of these findings in terms of associative processes not contingent on punishment—such as recency or stimulus impact (salience) effects stemming from the conjunction of the label with the termination of the buzzer at the end of each trial. It cannot be said, for example, that all children would be generally predisposed to imitate the experimenter's labels, for reasons unrelated to their anxiety-reducing value, unless the labels had been made aversive stimuli through their association with onset of punishment. Few children in the control paradigm used the label "blue," even though it occurred at the termination of buzzer and training trial. Embedding the label in a punitive context is clearly a prerequisite of its adoption by the children.[6]

A comparison of the frequencies of occurrence of the blue and red labels, following the LTP paradigm, also indicates no general tendency of the children to imitate the experimenter's verbalization, and again points to the specific anxiety-reducing function of the blue label. Of course, the red label would not have been expected to acquire much value from its association with relatively neutral events. Its restricted use following nurturant socialization might be regarded as more surprising, however, since there it occurs in the context of warm verbal approval and physical affection. It may be that the reinforcing effects of the kind of nurturance given are not as discrete and well defined as the effects of anxiety reduction. But perhaps a more plausible explanation is that the motivation to produce an evaluative label associated with reward is difficult to establish in a learning situation so largely defined by punishment. In any case, the findings suggest that positive self-evaluation requires its own distinct reinforcing consequences.

MAINTENANCE

One of the remarkable characteristics of the internalized derivatives of social learning is their persistent maintenance in the absence of external surveillance and reinforcement. Aversive responses, in particular, can often be quite resistant to extinction (Miller, 1959; Solomon & Brush, 1956), and one would expect that self-criticism, a response that is even initially intrinsically reinforced, might easily become independent of its original conditions of learning. As has already been pointed out, the punishment and criticism of parents sometimes follow fairly rapidly upon occurrence of a response to which internalized anxiety is to be attached. Yet once the child's anxiety comes to precede its punishment, certain cue components of the punishment become associated with anxiety reduction. Correspondingly, in the experimental paradigm with labeling at termination of punishment, onset of punishment was immediately coincident with onset of the transgression signal, but the (external) presentation of the critical label directly coincided with the signal's termination. Such temporal relationships among the cues to which anxiety is attached, the occurrence of punishment, and the stimulus properties of aversive responses, appear to contribute as much to the maintenance of these responses as to their acquisition (Brush et al., 1955; Church et al., 1956).

---

[6] The buzzer, though used as a cue, undoubtedly had some inherent aversive properties, the termination of which may have been sufficient in itself, for a few children, to lend anxiety-reducing value to the blue label in the control paradigm. The occasional use of the blue label by children exposed to the paradigms where it appeared directly upon transgression may be attributable to covert situational cues or irregularities of timing which permitted too much anticipatory anxiety before its occurrence.

Part of the relative insensitivity that self-criticism can show to changes in external conditions may be a function of the resistance to extinction of the intense anxiety attached to some transgressions (Bitterman & Holtzman, 1952; Solomon & Wynne, 1954). The actions around which parents hope to inculcate the most permanent and generalized internalization are those which they are likely to punish most severely. The failure of anxiety to disappear, when apparently deprived of its original support in the form of explicit punishment, may also follow from a kind of inconsistent reinforcement related to generalization of perceived punitive events. Children may perceive punitive implications even where punishment is not intended, because they can so easily assimilate both social and impersonal events to their experience of the wide spectrum of punitive capacities previously represented in the behavior of socializing agents. In the ecology of interaction between parents and children, inconsistency of perceived punishment would be inevitable. And even beyond childhood, the most infrequent application of real or perceived punishment would be sufficient to interfere with the extinction of internalized anxiety.

While internalized reactions to transgression do persist in the absence of predictable external punishment, their endurance when followed by punishment perhaps even more effectively shows that their reinforcement is not tied to external consequences. Children are often punished for their actions despite the intervening implicit or overt presence of self-criticism. The cue properties of their self-critical responses would, under these circumstances, precede external punishment. They would therefore no longer have quite the same signal value, with respect to onset and termination of anxiety, which they initially had as a component of punishment. The anxiety-reducing status of self-criticism could not be sustained under such conditions, if it actually remained dependent on the temporal relationship of the critical labels to the external events which originally marked anxiety reduction. Moreover, the continued experience of social punishment affords children many opportunities to observe that internalized anxiety can be terminated without their criticizing or punishing themselves, though not in a way that is equally under their control. It is possible, of course, that the irregular presence of punishment serves primarily to strengthen the anxiety that motivates self-criticism. A number of studies with animals have indicated that punishment of aversive responses heightens their resistance to extinction (Gwinn, 1949; Sidman, Herrnstein, & Conrad, 1957; Solomon & Wynne, 1954). But whatever the effects of external punishment upon the self-critical response, once it is established, it seems necessary to attribute its persistent maintenance to the anxiety-reduction attached to its intrinsically produced cues.

In order to assess the durability of the tendency to reproduce a verbal component of punishment, three distinct attempts were made to eliminate the use of the blue label among additional groups of children, after they had acquired it through exposure to the socialization paradigm with labeling at termination of punishment. The three procedures may be designated as extinction paradigms, even though the original reinforcement (anxiety reduction) for the self-critical response cannot really be withdrawn by changes in external conditions. One group was taken through a procedure intended to diminish the anxiety-reducing properties of the blue label by changing its temporal relationship to both punishment and anxiety. The procedure used with the second group was directed to the question of whether the label would retain its value over a period during which punishment and anxiety reduction occurred without its being verbalized by either child or socializing agent. The third paradigm, which made the most direct attack upon the autonomy of self-criticism, removed punishment as well as verbalization of the critical label, in an attempt to extinguish the anxiety attached to the transgression.

Three Extinction Paradigms

Subjects

Thirty-seven additional subjects, drawn from the same population used in the original four socialization paradigms, were exposed to the LTP paradigm. Thirty of these new subjects met the previously employed acquisition criterion, and 10 were assigned to each of the three extinction series.

Procedure

In an effort to eliminate the self-critical response before it was given repeated overt verbalization, all of the extinction procedures were begun as soon as *S* verbalized the label "blue" on either of the two acquisition test trials (11 or 12), and all subsequent trials were run as the extinction series. The procedures were run so that each two transgression (buzzer) trials were followed by a nontransgression trial, in order to avoid uninterrupted punishment in the first two extinction paradigms. Thus, beginning with the first acquisition test trial, the sequence of transgression trials during extinction was: 11, 12, 14, 15, 17, 18–etc.[7]

Extinction Paradigm 1. In this series, *E* continued to use, on transgression trials, the same verbal eliciting stimuli used on acquisition test trials. As soon as *S* gave the self-critical response on either Trial 11 or 12, and regardless of *S*'s response on subsequent transgression trials, *E* rose and said, while walking to *S*'s table: "Well, you lose some of your Tootsie Rolls."[8] The *E* then removed two Tootsie Rolls from *S*'s supply and terminated the buzzer, but no longer used the label blue. On nontransgression trials, *E* simply rose and reset the levers on the machine (without using the label "red"). The effect of this procedure was that the label "blue," when now "presented" by *S* during transgression trials, was consistently followed by punishment and temporary continuation of the buzzer that signaled transgression. Extinction trials were run until there were four consecutive transgression trials without verbalization of the blue label, or until 10 transgression trials had occurred, although additional trials were permitted if *S* showed signs of extinguishing.

Extinction Paradigm 2. The blue label was eliminated from punishment during this procedure, and *E* also discontinued the verbal stimuli used to elicit *S*'s overt expression of the label. The *S* was consequently exposed to termination of both punishment and buzzer without verbalization of the label. Following *S*'s self-critical response on Trial 11 or 12, *E* rose, administered the usual punishment without using the blue label, and terminated the buzzer. On all subsequent transgression trials, *E* carried out the same procedure, beginning immediately upon onset of the buzzer. On nontransgression trials, *E* reset the levers without comment. The extinction series was run through Trial 25. Trials 26 and 27 were extinction test trials, and restored the verbal eliciting stimuli used by *E* on acquisition test trials. If *S* verbalized the blue label on neither of the two test trials, the self-critical response was considered to be extinguished.

---

[7] Children to be taken through the two extinction paradigms in which punishment was continued were provided, before training began, with 20 Tootsie Rolls (instead of the 15 provided in the original socialization paradigms), so that they could continue to be deprived of them during the entire sequence of transgression trials throughout the acquisition and extinction procedures.

[8] The expletive, "No!," used as a component of punishment during training trials, was removed in this first extinction paradigm. Its continued use, once the child began to verbalize the blue label, might have led to elimination of the label for reasons unrelated to the change in its temporal relationship to anxiety and punishment.

Extinction Paradigm 3. In this paradigm, both punishment and verbalization of the label "blue" were entirely removed. When *S* gave the self-critical response on Trial 11 or 12, *E* merely rose, walked over to *S*'s table, and terminated the buzzer. On subsequent extinction trials, both transgression and nontransgression, *E* continued to simply walk over and reset the levers without comment. The extinction criterion was *S*'s failure to use the self-critical label on either of the two extinction test trials (26 and 27), when *E*'s eliciting stimuli were restored.

Results

The effective induction of self-criticism, in the socialization paradigm with labeling at termination of punishment, was again confirmed by the fact that 30 of the 37 children taken through the paradigm met the acquisition criterion. Table 3 shows that, in each of the three extinction groups, only two of the ten children used met the extinction criterion. None of the children in whom the response failed to extinguish gave any indication of relinquishing their use of the critical label. Under the first extinction procedure, children who did not meet the extinction criterion used the label uniformly throughout the extinction trials. And in the second and third paradigms, children who retained the response invariably gave it on both of the extinction test trials.

The three extinction paradigms do not go to the limit of effort that could be made to eliminate the self-critical response. Other methods which have been used to facilitate extinction of aversive responses, such as pairing the anxiety-inducing stimulus with positive reinforcement (Hall, 1955) or rewarding an incompatible response (Solomon & Wynne, 1954), are not represented here. It seems apparent, however, that the experimentally socialized labeling response has some of the tenacity that we might expect of internalized self-evaluation.

SOME FURTHER THEORETICAL IMPLICATIONS

Reproduction of Punishment

The experimental literature on punishment learning typically centers on the conditioning of anxiety (or "aversiveness") to the organism's own response cues, and the consequences of punishment are indexed in the suppression of the punished behavior (Dinsmoor, 1954; Estes, 1944; Miller, 1959; Mowrer & Viek, 1948). The inhibition of transgressions would be, of course, one of the derivatives of their definition through social punishment. But self-criticism provides a response for reducing the anxiety that follows the commission of a transgression. The response does not require reinforcement by avoidance or termination of external punishment, and it actually reproduces stimulus events which have been previously part of that punishment. We are confronted, then, with an anxiety-reducing response whose acquisition is contingent on the prior association of its stimulus prop-

TABLE 3

FREQUENCY OF EXTINCTION AND NONEXTINCTION OF THE SELF-CRITICAL RESPONSE UNDER EACH OF THREE EXTINCTION PARADIGMS

Response	Extinction 1 ($N = 10$)	Extinction 2 ($N = 10$)	Extinction 3 ($N = 10$)
Extinguished	2	2	2
Did not extinguish	8	8	8

erties with the presence rather than with the absence of punishment.

It is not really paradoxical that self-criticism can be acquired, and often permanently maintained, despite the fact that it reproduces punishment. Aversive responses have frequently been shown to be at least as sensitive to anxiety reduction as they are to punishment avoidance, when these two outcomes are partially separated, and sometimes they seem relatively impervious to their punitive consequences (Gwinn, 1949; Sidman & Boren, 1957; Solomon & Wynne, 1954). It is possible, in fact, to interpret certain studies of both human and animal subjects (D'Amato & Gumenik, 1960; Farber, 1948; Maier, 1949) as indicating that, under high anticipatory anxiety and with no easily available avoidance response, subjects will move in the direction of external punishment because of its anxiety-attenuating significance. In a similar way, the anxiety that motivates self-criticism is initially dependent on the motivating properties of punishment, while the reinforcement of the response is dependent on its replication of the anxiety-reducing signal properties of punishment. Self-punitive responses are, therefore, clearly not to be regarded as having been originally learned as avoidance responses. Their self-reinforcing character is acquired on the basis of anxiety-reduction as the "primary reinforcement."

It is interesting to note that Freud (1936, Ch. 8) saw anxiety reduction as the function of children's adoption of the punitive or threatening behavior of a model, and that its role has also been previously suggested in concepts such as "identification with the aggressor" (A. Freud, 1946, Ch. 9) and "defensive identification" (Mowrer, 1950, Ch. 21). Certainly it is not difficult to perceive that the mechanism of reinforcement in the learning of self-criticism may serve as the general mechanism for what may be called identification through aversive learning—that is, for the acquisition of any response in which, without the benefit of external reinforcement, the individual directs toward himself the socially punitive behavior that he has previously experienced from without.[9]

Further research on the anxiety-reducing properties of social punishment, and on their imitative reproduction, might well give attention to the intensity of punishment. There is no reason to assume that the punishment to be reproduced must be less "painful" than the antecedent anxiety. Nevertheless, when punishment is very severe, whatever decrement of anticipatory anxiety attaches to its cue properties could be overshadowed by the increment of aversive stimulation from its drive properties. Experimental socialization paradigms might be used to compare the conditions required for the learning of self-criticism with those required for the self-directed replication of nonverbal components of punishment, some of which would be more inherently unpleasant than a verbal label. Such paradigms might also be devised so as to suggest the theoretical extensions necessary to account for the child's learning to direct its reproduction of punishment toward others. They could further be used to examine other relationships of timing and magnitude among components of punishment. Presumably, any component of punishment could assume anxiety-reducing value, provided that it follows some minimal internalized anxiety. But the breadth of the phenomenon may be limited by the duration and relative severity of the remaining punishment that follows a given component. Verbal social responses which already have become mildly punitive to the child are often used by socializing agents as cues to signal more severe punishment if a transgression is completed or continued. It would be interesting to know whether such cues attain some anxiety-reducing value, despite their position as warning signals, because of their mediation of inhibitory responses which avoid the punishment. Finally, perhaps the most intriguing experimental translations of naturalistic socialization would be in the area of inconsistent punishment. Studies of the discriminative events within which punishment is inconsistent, and of the extent to which socializing agents provide consequences of transgression other than direct punishment, might reveal that inhibition, self-punishment, and other internalized reactions to transgression have common elements other than the anxiety that motivates them and the anxiety reduction that serves as their reinforcement.

---

[9] Children also often react to the transgressions of others by reproducing the same punitive responses to which they have been previously subjected. A tentative beginning toward accounting for this phenomenon might be offered by noting that stimulus generalization may operate to cause the transgressions of others to arouse the child's anxiety, and to elicit its reproduction of punishment, if they are similar to those for which the child itself has been previously punished. One must also account, however, for the change in the target of the punishment.

Nurturance and the Internalized Consequences of Punishment

A comparison of the effects of the various socialization paradigms described earlier indicates that the internalization of punishment does not rest on a broader predisposition to reproduce the behavior of an otherwise nurturant or affectionate model. Nor does it appear, as some theorists have suggested (Kagan, 1958; Maccoby, 1959; Whiting, 1960), that self-criticism represents a generalized tendency to exercise control over the threatening punitive resources at the disposal of a socializing agent. While the duration and intensity of nurturant treatment are more extended in a parent-child relationship, the warm approval and physical affection given to the children in the nurturance paradigm would certainly have been expected to attach some positive value to the model's role and to cause punishment to be experienced as a withdrawal of affection. And, in view of the effectiveness of the paradigm with labeling at termination of punishment, the failure of the nurturance paradigm to induce many self-critical responses cannot be attributed simply to an insufficiency of nurturance. The reproduction of social punishment appears to be acquired only when the relevant components of punishment have a circumscribed temporal relationship to an anticipatory aversive state.

The contributions of punishment and anxiety to the learning of self-criticism should not be extrapolated, of course, to other kinds of imitative social learning, in which children do appear to provide their own reinforcement by reproducing those stimulus aspects of a model which have been directly associated with nurturance or reward. Persuasive theoretical cases have been made for the place of a reinforcement mechanism derived from reward in much of imitative behavior (Mowrer, 1950, Ch. 21; Mowrer, 1960b, Ch. 3; Sears et al., 1957, Ch. 10; Whiting & Child, 1953, Ch. 11); and there are empirical findings (Bandura & Huston, 1961; Mussen & Distler, 1959; Sears, 1953) which support the common observation that the warmth and affection of socializing agents foster the child's tendency to adopt a great variety of their responses. We can hardly be surprised, however, to find that the reproduction of social punishment cannot be subsumed under the consequences of a model's rewarding characteristics. Children sometimes reproduce behavioral pieces of the most entirely frightening figures which they encounter in both reality and fantasy. Similar phenomena in adults are described in Bettelheim's (1943) well-known report of life in a concentration camp, and are also seen by the sensitive observer under ordinary social circumstances. Conversely, there is no reason to expect the child's motivation to reproduce the behavior of a nurturant model to be so powerfully indiscriminate, even under the duress of withdrawal of affection, as to encompass the model's punishment. An account of the diversity of modeling responses usually taken as evidence of identification requires, then, an extension of learning processes derived from punishment as well as of those derived from reward.[10]

The internalized consequences of social punishment would, nevertheless, not be totally indifferent to the nurturance and affection of socializing agents. Numerous descriptive surveys of internalized responses to transgression, whose rough indices of internalization have given much attention to self-criticism, indicate the absence of any consistently direct relationship between extent of internalization in the child and the amount of nurturance given by parents (Bronfenbrenner, 1961; Heinecke, 1953; Sears et al., 1957, Ch. 10; Whiting & Child, 1953, Ch. 11). There is evidence, however, that the usual effects of punishment upon internalization require a minimally nurturant context. The children of extreme parental groups, who are highly rejecting or punitive toward their offspring, appear to show poor internalization in their social behavior (Bandura & Walters, 1959; Nye, 1958; Sears et al., 1957, Ch. 10; Whiting & Child, 1953, Ch. 11). Loss of affection is probably a major element in the effectiveness of any form of social punishment administered by parents to their children. This element could be negligible, if parents have been so generally cold or punitive as to adapt the child to a high level of aversive stimulation in their presence. Their specific punishments might then not produce an increment of anxiety sufficient to independently attach itself to the child's own response cues, so that it could motivate self-criticism or other reactions to transgression even in the absence of external control. It seems even more probable that parents who are generally lacking in nurturance use unusually severe punishments. The intensity of ordinary social punishment permits its components to have some significance for attenuating anxiety and is not likely to disrupt the child's attention to this significance. Severe punishment, in contrast, may evoke competing responses, such as withdrawal or escape, which are more closely associated with anxiety reduction than is any aspect of the punishment itself.

The key role that has been assigned to self-criticism in many surveys of internalization also affords an opportunity to inferentially confirm the importance of the temporal relationships so vividly illustrated in the experimental paradigms. A number of investigators (Allinsmith, 1960; Aronfreed, 1961; MacKinnon, 1938; Sears et al., 1957, Ch. 10; Whiting & Child, 1953, Ch. 11) have found a rather limited

[10] Even if the effects of the parents' nurturance were so powerful as to generalize and attach some value to their punitive presence, a gradient of generalization would yield the hardly credible prediction that, if the child transgressed and consequently felt deprived of their affection, it would then be more likely to apply their rewarding actions to itself than it would any component of their punishment.

positive association between the extent of children's internalization of responses to transgression and their parent's use of methods of punishment variously referred to as "love oriented," "psychological," or "induction" techniques. Withdrawal of affection has always been included among such techniques, and it often has been presumed to be their dominant common element. But the experimental findings reported here do not support the view that an explicit withdrawal of affection by a nurturant model induces a state of deprivation which is singularly effective in motivating a child's reproduction of the model's critical evaluative responses. The apparent discrepancy may be resolved by observing that, while loss of affection would enter into the child's experience of any form of punishment employed by its parents, their use of withdrawal of affection as the focus of their punishment might well incorporate certain features of administration and timing which are much less characteristic of other disciplinary techniques. These features are precisely those which have been artificially removed in the experimental nurturance paradigm, where the socializing agent's critical labels acquire little or no anxiety-reducing value because they are presented immediately upon the child's awareness of transgression, before substantial anxiety can be evoked.

Parents who focus on withdrawal of affection in their discipline may be those who are also most oriented toward inducing an internal governor of conduct in their children, in contrast to other parents who are more oriented toward sensitizing their children to the punitive external consequences of transgressions with which they are visibly confronted (Aronfreed, 1961; Bronfenbrenner, 1958). Accordingly, parents who emphasize withdrawal of affection may be more disposed to punish transgressions well beyond the point of their occurrence. And they may tend to maintain the cognitive and affective salience of their children's transgressions by reinstating affection only after some time has elapsed. It is the manner in which they use withdrawal of affection, then, that would induce in their children internalized anxiety of greater and more uncertain duration than that induced by other more direct and quickly terminated punishments. This more variable and prolonged period of anxiety would sustain a greater probability that the child would reproduce components of parental criticism, or otherwise make its own anxiety-reducing responses, rather than wait upon external events. The fact that deprivation of affection does not require the physical presence of the parents (and is often effected by their absence) is also useful in understanding why it makes the child prone to produce anxiety-reducing stimuli through its own resources.[11]

Cognitive Processes in Self-evaluation

The critical label "blue," used in the experimental socialization paradigms, was primarily a color designation in the children's previous experience. It was applied to a transgression that was not really within their volitional control and not clearly related to evaluative standards with which they were already familiar. The presence and absence of the label during training were not even respectively associated with responses which could be intrinsically identified by the children. Yet, under the appropriate temporal conditions, they adopted the blue label to refer to their own actions—a behavioral change that appears to demonstrate that the motivational prerequisites for the learning of self-criticism may be present in the most limited cognitive surroundings. The infrequent application of the red label to nonpunished actions, even among children who did use the blue label and who would surely have some cognition of the blue-red distinction, suggests further that these motivational precursors are essential to the acquisition of self-evaluative responses.

---

[11] The survey findings are often ambiguous because highly verbal parental behaviors, such as reasoning, explanation, and appeals to the child's already established standards, are commonly grouped together with actual withdrawal of affection within the class of disciplinary techniques which seem to have some association with greater evidence of internalization in the child. The findings could therefore indicate that it is the verbal and cognitive context of punishment, rather than the specific properties of deprivation of affection as a method of punishment, that fosters the internalization of self-criticism.

That the cognitive context of a labeling response is insufficient to fully define its evaluative properties, without taking into account the affective conditions under which it is acquired, is also made clear by more general considerations. The discrepancies which individuals may perceive between their behavior and their standards, particularly in areas of performance or achievement, are subjected to evaluative cognition and sometimes result in subsequently rewarded modifications of behavior. But their evaluative processes would not require that they reproduce, and apply to their own behavior, a verbal component of the punishment to which they have been previously exposed. Nor do the rewarding consequences of their behavioral corrections reveal the motivation or the reinforcement for their self-critical responses. When self-criticism does occur, its reinforcement cannot be understood without reference to its aversive motivational antecedents, which are in turn dependent on the affective correlates of evaluative processes. We seem to be in the habit of regarding self-evaluation as being the instrumental verbal responses which are motivated by these affective correlates. But evaluative processes might better be conceived of as cognitive and affective phenomena which do not in themselves require verbal praise or censure.

The treatment of self-criticism as an instrumental labeling response does not circumvent the importance of understanding the complex cognitive processes upon which self-evaluation is often based. We might consider, for example, whether reproducing the verbal components of social punishment is to be regarded as a moral response—or, to put it another way, whether it is to be taken as evidence of guilt. In the normative framework of our society, the term *moral* is commonly restricted to actions which follow from some degree of judgment exercised by the actor. The evaluative dimensions of such judgment are presumed to be ethical in nature, in the sense that they make articulate distinctions between right and wrong or good and bad. But many of the internalized derivatives of social reinforcement, including the application of either punitive or rewarding labels to one's own actions, may require only very rudimentary and tenuous cognitive support. Many other internalized responses reflect socially derived cognitive processes which are highly developed along evaluative dimensions which are not necessarily moral. This would surely be the case with the self-evaluation that attaches to success or failure in certain kinds of performance or control.

There are formidable problems associated with any attempt to define moral cognition in a way that escapes cultural relativism. It might be possible, however, to assign invariant properties to moral evaluation by reference to particular kinds of cognitive processes. Certainly the moral properties of an evaluative response would be determined by its cognitive dimensions. In western society at least, moral cognition appears to be characterized by reference to the consequences of an act for others, and, to a less definitive extent, by reference to the intention of the act. The labeling responses induced by the socialization paradigms reported here cannot be regarded as moral responses, since they would rest in only the most primitive way on cognitive mediating processes. But the experimental phenomenon does demonstrate learning and motivational mechanisms which are fundamental to self-criticism, regardless of its moral status. With appropriate cognitive buffering across a sufficient variety of situations, the label "blue" could be endowed with an unmistakable moral connotation.[12]

These suggested limitations on the boundaries of morality also imply that the concept of guilt is too liberally employed when it is used to describe either the arousal of anxiety in the absence of external punishment or the presence of a self-critical labeling response that reduces the anxiety. The concept might be reserved more appropriately for an affective state with certain distinctive cognitive features. Both introspection and common usage suggest that the experience of guilt is compounded of the anxiety aroused by transgression and the evaluative perceptions which come through the filter of moral cognition. While this cognition can be normatively circumscribed for a given society and operationalized in verbal behavior, guilt must be viewed as an essentially phenomenological construct of variable substance.

---

[12] An excellent discussion of the place of moral cognition within the larger realm of internalized responses can be found in some of Kohlberg's (1963) recent work.

REFERENCES

ALLINSMITH, W. The learning of moral standards. In D. R. Miller, G. E. Swanson, et al., *Inner conflict and defense.* New York: Holt, 1960. Pp. 141–176.

ARONFREED, J. Moral behavior and sex identity. In D. R. Miller, G. E. Swanson, et al., *Inner conflict and defense.* New York: Holt, 1960. Pp. 177–193.

ARONFREED, J. The nature, variety, and social patterning of moral responses to transgression. *J. abnorm. soc. Psychol.,* 1961, **63**, 223–240.

ARONFREED, J. The effects of experimental socialization paradigms upon two moral responses to transgression. *J. abnorm. soc. Psychol.,* 1963, **66**, 437–448.

ARONFREED, J., CUTICK, R. A., & FAGAN, S. A. Cognitive structure, punishment, and nurturance in the experimental induction of self-criticism. *Child Developm.,* 1963, **34**, 281–294.

BANDURA, A., & HUSTON, ALETHA C. Identification as a process of incidental learning. *J. abnorm. soc. Psychol.,* 1961, **63**, 311–318.

BANDURA, A., & WALTERS, R. H. *Adolescent aggression.* New York: Ronald Press, 1959.

BECK, R. C. On secondary reinforcement and shock termination. *Psychol. Bull.,* 1961, **58**, 28–45.

BETTELHEIM, B. Individual and mass behavior in extreme situations. *J. abnorm. soc. Psychol.,* 1943, **38**, 417–452.

BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.,* 1952, **47**, 615–623.

BRADY, J. V., SCHREINER, L., GELLER, I., & KLING, A. Subcortical mechanisms in emotional behavior: The effect of rhinencephalic injury upon the acquisition and retention of a conditioned avoidance response in cats. *J. comp. physiol. Psychol.,* 1954, **47**, 179–186.

BRONFENBRENNER, U. Socialization and social class through time and space. In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (3rd ed.) New York: Holt, 1958. Pp. 400–425.

BRONFENBRENNER, U. Freudian theories of identification and their derivatives. *Child Developm.,* 1960, **31**, 15–40.

BRONFENBRENNER, U. Some familial antecedents of responsibility and leadership in adolescents. In L. Petrullo & B. L. Bass (Eds.), *Leadership and interpersonal be-*

havior. New York: Holt, Rinehart, & Winston, 1961. Pp. 239–272.

Brush, F. R., Brush, Elinor, S., & Solomon, R. L. Traumatic avoidance learning: The effects of CS–US interval with a delayed-conditioning procedure. *J. comp. physiol. Psychol.*, 1955, **48**, 285–293.

Buchanan, G. The effects of various punishment-escape events upon subsequent choice behavior of rats. *J. comp. physiol. Psychol.*, 1958, **51**, 355–362.

Church, R. M. Transmission of learned behavior between rats. *J. abnorm. soc. Psychol.*, 1957, **54**, 163–165.

Church, R. M., Brush, F. R., & Solomon, R. L. Traumatic avoidance learning: The effects of CS–US interval with a delayed-conditioning procedure in a free-responding situation. *J. comp. physiol. Psychol.*, 1956, **49**, 301–308.

D'Amato, M. R., & Gumenik, W. E. Some effects of immediate versus randomly delayed shock on an instrumental response and cognitive processes. *J. abnorm. soc. Psychol.*, 1960, **60**, 64–67.

Darby, C. L., & Riopelle, A. J. Observational learning in the rhesus monkey. *J. comp. physiol. Psychol.*, 1959, **52**, 94–98.

Davitz, J. R., Mason, D. J., Mowrer, O. H., & Viek, P. Conditioning of fear: A function of the delay of reinforcement. *Amer. J. Psychol.*, 1957, **70**, 69–74.

Dinsmoor, J. A. Punishment. 1. The avoidance hypothesis. *Psychol. Rev.*, 1954, **61**, 34–46.

Estes, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **574** (3, Whole No. 263).

Farber, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.

Freud, Anna. *The ego and the mechanisms of defense.* New York: International Universities Press, 1946.

Freud, S. *The problem of anxiety.* New York: Norton, 1936.

Gleitman, H. Place learning without prior performance. *J. comp. physiol. Psychol.*, 1955, **48**, 77–79.

Goodson, F., & Brownstein, A. Secondary reinforcing and motivating properties of stimuli contiguous with shock onset and termination. *J. comp. physiol. Psychol.*, 1955, **48**, 381–386.

Gwinn, G. T. The effects of punishment on acts motivated by fear. *J. exp. Psychol.*, 1949, **39**, 260–269.

Hall, Julia C. Some conditions of anxiety extinction. *J. abnorm. soc. Psychol.*, 1955, **51**, 126–132.

Hartshorne, H., & May, M. A. Studies in the nature of character. Vol. 1. *Studies in deceit.* New York: Macmillan, 1928.

Heinecke, C. M. Some antecedents and correlates of guilt and fear in young boys. Unpublished doctoral dissertation, Harvard University, 1953.

Hill, W. F. Learning theory and the acquisition of values. *Psychol. Rev.*, 1960, **67**, 317–331.

Jenkins, W. O. A temporal gradient of derived reinforcement. *Amer. J. Psychol.*, 1950, **63**, 237–243.

Kagan, J. The concept of identification. *Psychol. Rev.*, 1958, **65**, 296–305.

Kamin, L. J. The delay-of-punishment gradient. *J. comp. physiol. Psychol.*, 1959, **52**, 434–436.

Kohlberg, L. Moral development. In M. L. Hoffman (Ed.), *Annu. Rev. Child Developm. Res.*, **1**, 1964, in press.

Littman, R., & Wade, E. A. A negative test of the drive-reduction hypothesis. *Quart. J. exp. Psychol.*, 1955, **7**, 56–66.

Maccoby, Eleanor E. Role-taking in childhood and its consequences for social learning. *Child Developm.*, 1959, **30**, 239–252.

MacKinnon, D. W. Violation of prohibitions. In H. A. Murray et al. (Eds.), *Explorations in personality.* New York: Oxford Univer. Press, 1938. Pp. 491–501.

Maier, N. R. F. *Frustration: The study of behavior without a goal.* New York: Holt, 1949.

Miller, N. E. Liberalization of basic S–R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959. Pp. 196–292.

Miller, R. D., Swanson, G. E. et al. *Inner conflict and defense.* New York: Holt, 1960.

Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

Montgomery, K. C., & Galton, B. B. A test of the drive-reduction explanation of learned fear. Paper read at Eastern Psychological Association, Philadelphia, April 1955.

Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.

Mowrer, O. H. *Learning theory and behavior.* New York: Wiley, 1960. (a)

MOWRER, O. H. *Learning theory and the symbolic processes.* New York: Wiley, 1960. (b)

MOWRER, O. H., & AIKEN, E. G. Contiguity vs. drive-reduction in conditioned fear: Temporal variations in conditioned and unconditioned stimulus. *Amer. J. Psychol.,* 1954, **67**, 26–38.

MOWRER, O. H., & LAMOREAUX, R. R. Avoidance conditioning and signal duration—a study of secondary motivation and reward. *Psychol. Monogr.,* 1942, **54**(5, Whole No. 247).

MOWRER, O. H., & VIEK, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.,* 1948, **43**, 193–200.

MUSSEN, P., & DISTLER, L. M. Masculinity, identification, and father-son relationships. *J. abnorm. soc. Psychol.,* 1959, **59**, 350–356.

NEFZGER, W. D. The properties of stimuli associated with shock reduction. *J. exp. Psychol.,* 1957, **53**, 184–188.

NYE, I. F. *Family relationships and delinquent behavior.* New York: Wiley, 1958.

PIAGET, J. *The moral judgment of the child.* Glencoe, Ill.: Free Press, 1948.

SCHOENFELD, W. N., ANTONITIS, J. J., & BERSH, P. J. A preliminary study of training conditions necessary for secondary reinforcement. *J. exp. Psychol.,* 1950, **40**, 40–45.

SEARS, PAULINE S. Child-rearing factors related to playing of sex-typed roles. *Amer. Psychologist,* 1953, **8**, 431. (Abstract)

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child-rearing.* Evanston, Ill.: Row, Peterson, 1957.

SIDMAN, M., & BOREN, J. The relative aversiveness of warning signal and shock in an avoidance situation. *J. abnorm. soc. Psychol.,* 1957, **55**, 339–344.

SIDMAN, M., HERRNSTEIN, R. J., & CONRAD, D. G. Maintenance of avoidance behavior by unavoidable shocks. *J. comp. physiol. Psychol.,* 1957, **50**, 553–557.

SOLOMON, R. L., & BRUSH, ELINOR S. Experimentally derived conceptions of anxiety and aversion. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1956.* Lincoln: Univer. Nebraska Press, 1956. Pp. 212–305.

SOLOMON, R. L., & TURNER, LUCILLE H. Discriminative classical conditioning in dogs paralyzed by curare can later control discriminative avoidance responses in the normal state. *Psychol. Rev.,* 1962, **69**, 202–219.

SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychol. Rev.,* 1954, **61**, 353–385.

WHITING, J. W. M. Sorcery, sin, and the superego: Some cross-cultural mechanisms of social control. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1959.* Lincoln: Univer. Nebraska Press, 1959. Pp. 174–195.

WHITING, J. W. M. Resource mediation and learning by identification. In I. Iscoe & H. W. Stevenson (Eds.), *Personality development in children.* Austin, Tex.: Univer. Texas Press, 1960. Pp. 112–126.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality.* New Haven: Yale Univer. Press, 1953.

(Received August 17, 1962)

Psychological Review
1964, Vol. 71, No. 3, 219–237

INTRATRIAL AND INTERTRIAL RETENTION: NOTES TOWARDS A THEORY OF FREE RECALL VERBAL LEARNING[1]

ENDEL TULVING

University of Toronto

Trial-to-trial analysis of recall of individual items in a multitrial free recall learning experiment shows that the traditional learning curve can be expressed as an additive function of intratrial and intertrial retention. Intertrial retention, measured in terms of number of items retained from one trial to the next, increases as a logarithmic function of trials, while intratrial retention, measured in terms of number of items retained from the input phase of a given trial to the output phase of that trial, remains practically invariant over trials. Intertrial retention is positively correlated with subjective organization, giving support to the hypothesis that the increase in intertrial retention as a function of practice reflects the growth in the size, but not necessarily the number, of subjective units of material that the subject can retrieve from the memory storage.

One of the most universally known and reliable phenomena of memory lies in the improvement in recall of verbal materials under conditions of practice. Experiments in verbal learning, as in many other areas of psychological inquiry, often yield inconsistent and contradictory data, but no experimenter has ever reported that his human subjects, appropriately instructed, failed to increase their recall over trials of an experiment.

In spite of, or perhaps because of, the wide generality of the phenomenon, empirical and conceptual analyses of it have seldom been reported in the literature. A great deal of experimental evidence and attendant theoretical speculation is available about conditions and variables that influence the rate at which recall changes over trials, that is, about interaction effects involving trials as one of the independent variables, but systematic attempts to account for the effects of trials per se have not been popular.

The present paper reports an analysis of the trial-by-trial memorization of a list of words under the method of free recall (FR). In a typical FR experiment with which we are concerned the subject is exposed to a list of words in the "input" phase of a trial and is then asked to recall as many of these words as he can, in any order he wishes. Then the same words are presented again, usually in a different order, and again at the end of the presentation, in the "output" phase of the trial, the subject recalls the words he remembers. This procedure, alternating input and output phases, can be repeated for any number of cycles or trials.

The subject's task in the FR experiment appears to be relatively simple. Response learning in the sense of response integration (Mandler, 1954) is minimized whenever more or less familiar words are used, and associative learning is not necessary since the experimenter gives credit to the subject for words recalled regardless of their order. Memorization of a list of words in the FR experiment is rather reminiscent of that aspect of response learning that has to do with strengthening of responses specified by the input list so that they have "greater response strength than the many other responses in the repertoire which are not in the list the subject is to learn [Underwood & Schulz, 1960, p. 93]." If this reasoning is correct, then FR learning must involve only some of the processes underlying acquisition in serial and paired-associated learning experiments. It thus constitutes a simpler task for the subject to master and a simpler situation for the experimenter to analyze.

---

[1] This research has been supported by the National Research Council of Canada, under Grant No. APA-39, and by the National Science Foundation, under Grant No. GB-810. A preliminary version of the paper was read as an invited paper at the annual meeting of the Canadian Psychological Association in Montreal, June, 1961. The writer is indebted to George Mandler for constructive comments and cogent criticisms.

Probably because of its apparent simplicity, free-recall learning has failed to arouse much curiosity among experimenters. Only a handful of studies having to do with multitrial FR learning have been reported in the literature (Bruner, Miller, & Zimmerman, 1955; Horowitz, 1961; Murdock, 1960; Waugh, 1961). But it has appealed to several theorists who have constructed various stochastic models to describe its data (Bush & Mosteller, 1955; Miller & McGill, 1952; Waugh & Smith, 1962). All of these models have succeeded in providing good or even excellent mathematical descriptions of certain quantitative aspects of empirical findings. Yet they have not been primarily concerned with isolation and identification of processes involved in memorization of verbal material. The Waugh and Smith (1962) model perhaps goes farthest of the three in that it refers to three hypothetical processes—"labelling," "selecting," and "fixing"—identified with the three parameters of the model, but these processes are not related to any "classical psychological functions [p. 142]."

Another characteristic that the stochastic models have in common lies in their assumption that individual list items are memorized independently of one another. The recall of a given list item is supposed to have no effect on the recall or nonrecall of any other item. The fact that such an assumption is not tenable is well known to the authors of the statistical models, although the implications of the correct assumption of item relatedness have not yet been fully explored. In considering the problem briefly, Miller and McGill (1952) suggested that "associative clustering should affect the variability, not the rate, of memorization [p. 390]," but this somewhat startling hypothesis has not been subjected to an empirical test.

Finally, the description of FR learning provided by the stochastic models is difficult to reconcile with the fundamental fact that an individual list item, if it has been previously integrated, can be "learned" by the subject on one trial of a very short duration, in the sense that the subject can always recall it immediately after seeing or hearing it once. What is it that permits the single item to be labeled, selected, and fixed with a probability of unity, to use Waugh and Smith's language, if it is presented alone, but probabilities much smaller than unity when the item appears in a list? This particular criticism, of course, can be directed at almost any theoretical account of memorization, both formal and informal, that has ever been proposed, but the commonality of an error does not justify it.

Our present analysis of FR learning takes as its starting point the observation that a small unit of verbal material, such as an individual list word, is always "learned" at the time of its presentation to the subject. It then proceeds to examine to what extent these individual items are retained or forgotten, both within a single trial and over successive trials. The traditional learning curve will be described as a composite of intratrial and intertrial retention curves. Finally, intertrial retention as the main component of the subject's total recall performance will be shown to be related to the extent to which subjects organize their recall. It will be argued that the trial-by-trial improvement in recall is a consequence of the development of higher-order units of material which mediate the retrieval of the information from the memory storage.

ONE-TRIAL LEARNING AND INTRATRIAL FORGETTING OF INDIVIDUAL ITEMS

Let us begin by considering what happens on the first trial in a typical FR experiment. Each subject will be able to recall some but not all words in the output phase of the first trial. Such limitation of first-trial recall has traditionally been ascribed to some limitation in the learning mechanism, that is, it has been interpreted as reflecting the fact that the subject, for some reason or other, cannot *learn* all words on the first trial. For instance, in the well-known debate on whether associations in paired-associate learning tasks are formed in an all-or-none or incremental fashion, the participants, although disagreeing on the main issue, seem to agree on interpreting nonrecall of some response items as evidence for absence of, or incomplete formation of associations in the input phase preceding the recall test (Estes, 1960, 1961; Postman, 1963; Rock, 1957; Underwood & Keppel, 1962).

Such an interpretation, if taken literally, may prove somewhat misleading in pursuing the problems involved in verbal learning. Recent evidence shows clearly that a small unit of well-integrated verbal material, or an association between two such units, can practically always be recalled immediately following its presentation (Brown, 1958; Murdock, 1961a, 1961b, 1963; Peterson, Peterson, & Miller, 1961; Peterson, Saltzman, Hillner, & Land, 1962; Tulving & Arbuckle, 1963). When we use probability of response as a measure of learning, therefore, we must conclude that learning of a small unit of material is always complete on a single trial. Learning, in this sense, is neither incremental nor all-or-none, it is always "all."[2] Whether the small unit is presented alone or in a series of other units of the same class is immaterial as far as its one-trial acquisition is concerned. If we presented a list of words to the subject and warned him in advance that we might stop the presentation at any time and test him on the word seen last, we would have every reason to believe that his recall would be practically perfect.[3]

---

[2] Rock, in considering the problem of "why, in multiple-item learning situations, it is not possible as a rule to learn more than a few associations on any one trial," mentioned the possibility that an association is always formed and "that the failure to get all items right on a test following the entire series is a matter of forgetting of already formed associations [Rock, 1957, p. 192]." He rejected this interpretation for reasons that are not entirely clear to this writer.

[3] It is possible, of course, that subjects in certain types of experiments may in effect "ignore" some input items at the time of their presentation and concentrate their attention on a few specific items. One such selective strategy has been proposed by Feigenbaum and Simon (1962) in their account of the serial position curve. These strategies might indeed explain why the immediate recall of single items or associations in some experiments has been somewhat less than perfect, but their existence cannot change

This "learning" of an individual item—whether we call it learning, perception, fixation, registration, or what not, does not matter, as long as we know the operational referent of the term—depends solely upon variables operating prior to and at the time of the presentation of the item. It must be distinguished from the retention of the item which depends on the conditions prevailing at the time of the learning as well as on certain variables operating in the retention interval, however short. Thus the fact that an individual item is always "learned" following a single presentation does not mean that it might not be forgotten later. Small amounts of material, well within the classical memory span of the individual, often become unavailable for recall in a matter of a few seconds (Brown, 1958; Murdock, 1961a, 1961b, 1963; Peterson & Peterson, 1959; Tulving & Arbuckle, 1963). Nor does the conception of one-trial learning in any way preclude the possibility that additional presentations or rehearsal of the once learned and recalled item may change its resistance to forgetting. Just as overlearning of larger units of material, such as lists of nonsense syllables, has been long known to improve the retention of material (Ebbinghaus, 1885), smaller units have also been shown to benefit from repeated presentation (Peterson & Peterson, 1959; Peterson, Saltzman, Hillner, & Land, 1962; Waugh, 1962b).

It seems reasonable to argue, in view of the foregoing considerations, that limited recall on the first trial does not necessarily reflect incomplete learning, but rather, incomplete retention. Although the intratrial retention intervals, between the presentation and the attempted recall of each individual item, are short, many learned items can be forgotten within that interval. Whether such intratrial forgetting is interpreted in terms of the decay of memory traces (Broadbent, 1958; Brown, 1958; Conrad & Hille, 1958), in terms of input and output interference (Tulving & Arbuckle, 1963), or in terms of yet other factors, is not immediately relevant to our present purposes. What is important is the empirical fact that some individual items, although always recallable immediately following their presentation, do become unavailable for recall well within the interval occupied by a single trial. The recall scores on the first trial, therefore, reflect the combined effects of one-trial learning and intratrial forgetting, or simply of intratrial retention.

We may, if we wish, regard the subject's recall score on the first trial as a measure of his "immediate memory," provided that we realize that such "immediate" memory is limited for the simple reason that it is not immediate. In the FR experiment, it is only the last item that the subject could recall immediately after its presentation, but in a learning task in which the subject has to recall the items in the order of their presentation, recall is not immediate for any single item.

The notions of one-trial learning and intratrial forgetting suggest that it is not the fact of "storage" of list items that is at issue in the FR experiment, but rather the form of storage, or accessibility, of items (cf. Miller, Galanter, & Pribram, 1960). The list items have been "stored" in the subject's memory a long time before he appears for the experiment and the input list serves simply as a set of instructions as to which of the stored items the subject has to retrieve. In

---

the argument that one-trial learning can, and usually does, occur.

this sense, then, what the subject has to do in the FR experiment is to increase the accessibility of those items that are specified by the input list so that they can be readily retrieved from the storage.

Be it as it may, there are certain advantages in thinking of the first-trial recall as a measure of retention rather than learning. First, it might be easier to deal with questions such as why some items are retained and others forgotten within a single trial (e.g., Tulving & Arbuckle, 1963) than questions such as why some items are and others are not learned, or why each item is learned only to a certain extent and not completely. In terms of our existing conceptual tools as well as available empirical evidence we are probably better equipped to handle problems of retention and forgetting than of original acquisition. Second, and more important for the task at hand, the concept of intratrial retention leads naturally to the concept of intertrial retention and thus suggests a potentially useful way of looking at trial-by-trial recall data. The relation of intratrial and intertrial retention to the overall recall performance will become clear if we consider the pattern of recall of individual list items from trial to trial. This will be done in the next section.

TRIAL-TO-TRIAL ANALYSIS OF RECALL

We have argued that under appropriate conditions each list item is always learned on the first trial and that the subject's performance in the output phase of that trial is to be regarded as a measure of intratrial retention. The subject's performance on the second trial, and on all succeeding trials, can similarly be regarded more profitably as a measure of retention. On these trials, however, there are two sources of retained items. Some items could be recalled in the output phase of the second trial, for instance, even without the benefit of their re-exposure in the input phase of the second trial: the subject remembers some of the responses he made in the output phase of the first trial. The number of these responses provides a measure of intertrial retention. In addition to these responses, the subject, in the output phase of the second trial, also remembers some of the responses he made in the input phase of the second trial. Their number provides a measure of intratrial retention. The subject's overall recall performance on all trials except the first, therefore, consists of two components—intertrial and intratrial retention.

The conception of recall performance in terms of the process of retention or, looking at the other side of the coin, of forgetting, is not a novel one. McGeoch, for instance, more than 20 years ago, pointed out the relations involved:

> The changes in behavior (verbal responses) acquired during the first trial are retained, at least in part, until the second trial. There new ones are added to those retained; some or all of the results of the first and second trial are retained until the third trial, when more are added, and so on, until practice stops. . . . Not all of the acquisitions at each successive trial are carried over to the next; some are forgotten and must be refixated. A curve of learning represents a progressively greater balance in favor of retention, so that it is, in part, a retention curve. . . . Fixation and retention thus mutually interact in the course of what we call learning [McGeoch, 1942, pp. 4–5].

A single practice trial, as well as a series of trials, thus serves at least two functions. It provides the learner with the opportunity to study and retain individual items, but certain events occurring on the trial must also be responsible for forgetting of some other items. The traditional procedure used in verbal learning experiments, in which the experimenter assumes some kind of an equivalence of different list items and expresses the subject's performance in terms of the total number or proportion of items recalled, perhaps simplifies computations, but it clouds the two roles played by the trial.

The traditional method for describing the subject's performance in a verbal learning experiment is based on what might be called the trial-by-trial analysis of recall. On every trial, the experimenter partitions the total set of L items in the list into two mutually exclusive subsets, P_n and N_n. P_n (Performance) refers to the number of items that the subject has recalled on Trial n, while N_n refers to the number of items that the subject has failed to recall on that trial. $P_n + N_n = L$, hence the specification of P_n completely determines its complement N_n and nothing further can be gained from measuring N_n. The traditional learning curve is simply the plot of the size of the subset P_n against trials.

The trial-by-trial analysis of recall is to be contrasted with what will here be referred to as the trial-to-trial (TTT) analysis. This analysis is an extension of that used in Estes' (1960) "miniature experiments" to the typical multitrial experiment. In the TTT analysis, the Set L is partitioned with respect to pairs, triplets, etc., of successive trials. In the present paper, however, we shall be concerned with the TTT analysis as applied only to successive pairs of trials.

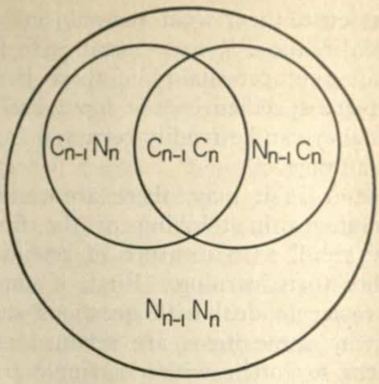


FIG. 1. Venn diagram representing four components of performance derived from the application of the trial-to-trial analysis to consecutive pairs of trials.

The subject's recall on two successive trials, $n - 1$ and n, defines within the Set L two subsets, P_{n-1} and P_n, thus placing each element of the Set L (each word in the list) into one of four mutually exclusive subsets. These four subsets are shown in the Venn diagram in Figure 1. The large circle represents the Set L, the small circle on the left represents the subset P_{n-1}, consisting of items that the subject recalled on Trial $n - 1$ and the small circle on the right represents subset P_n, consisting of items that the subject recalled on Trial n.

The four mutually exclusive subsets shown in the diagram are designated as follows:

$C_{n-1}C_n$ (or simply CC) is the intersection of the subsets P_{n-1} and P_n. It consists of items that occur in the subject's recall both on Trial $n - 1$ and on Trial n. The size of this subset can be regarded as an estimate of intertrial retention.

$N_{n-1}C_n$, (NC), consists of items that occur in the subject's recall on Trial n, but not on Trial $n - 1$. Its size provides an estimate of intratrial retention.

$C_{n-1}N_n$, (CN), consists of items recalled on Trial $n - 1$, but not recalled ("forgotten") on Trial n. Its size can be thought of as an estimate of intertrial forgetting.

$N_{n-1}N_n$, (NN), consists of items that the subject fails to recall both on Trial $n - 1$ and on Trial n. Its size can be regarded as an estimate of intratrial forgetting.

The logical definition of the four

subsets is straightforward, but the identification of the subsets with psychological processes of intratrial and intertrial retention and forgetting involves assumptions that may not be quite correct. For instance, referring to the *CC* component of performance as an estimate of intertrial retention ignores the possibility that an item in Subset *CC* was recalled on Trial *n*, forgotten by the time it was shown to the subject again in the input phase of Trial $n + 1$, "relearned" then, and retained until the output phase of Trial $n + 1$. Thus, both intertrial and intratrial retention may contribute items to the *CC* component which therefore may overestimate the amount of material that the subject can retain from one trial to the next without any intervening input. For the same reason the *NC* component may underestimate intratrial retention. Moreover, the extent of these errors of estimation may vary systematically over trials. These relations clearly constitute problems for further research. For the time being, however, it does not seem completely unreasonable to regard the four measures—*CC, NC, CN,* and *NN*—as estimates of intertrial and intratrial retention and forgetting as designated above.

If the TTT analysis is extended over all consecutive pairs of trials (Trials 0 and 1, 1 and 2, 2 and 3, . . . $N - 1$ and *N*, where *N* refers to the total number of trials), and if the size of each of the four subsets is plotted against trials, four curves are obtained: an intertrial retention curve, an intratrial retention curve, an intertrial forgetting curve, and an intratrial forgetting curve.

These four curves contain all the information that is available in the trial-by-trial "learning curve," but they also lay bare the anatomy of such a curve. Any two of the four curves can be mathematically derived from the other two, even though for a given pair of trials only one of the four measures is redundant with the other three. This fact reflects the complementarity of the definitions of the two processes, retention and forgetting: whatever the subject does not retain, he has forgotten, and vice versa. It also means that the traditional learning curve, in which *P* is plotted against trials, can be described in terms of any two of the four components.

In this paper, we shall be mainly concerned with the following relation:

$$P_n = C_{n-1}C_n + N_{n-1}C_n \qquad [1]$$

Equation 1 is a simple mathematical statement to the effect that performance is an additive function of two components, *CC* (intertrial retention) and *NC* (intratrial retention). Performance, either on a single trial or over a number of successive trials, can be analyzed into the *CC* and *NC* components, and these two components, again either on a single trial or over successive trials, can be synthesized into overall performance.

EMPIRICAL SYNTHESIS AND MATHEMATICAL ANALYSIS OF THE FREE RECALL LEARNING CURVE

Let us next examine some data from a simple FR experiment and demonstrate, first, how the learning curve can be derived from the empirically determined measures of *CC* and *NC*, and second, how the two components can be estimated from the overall performance data.[4]

Thirty-two summer school students served as subjects in the experiment. Each subject learned a list of 22 words on 22 trials by the method of FR. Words were presented to the subject by means of a Gerbrands memory drum, at the rate of one word per second. Subjects recalled words by speaking them out aloud. Their responses were recorded on tape and later transcribed. No definite amount of time was given for recall in the output phase; when the subject had been silent for 10 seconds, the input phase of the next trial was begun.

---

[4] Thanks are due to Albert S. Bregman who collected these data and wrote the electronic computer programs for analyzing them.

There were four different, but formally equivalent stimulus lists. Each list was learned by eight different subjects. The lists are shown in Table 1. The words were selected from the Thorndike-Lorge (1944) word book, according to the following criteria: (*a*) All words were two-syllable nouns, or words that could be used as nouns. (*b*) No two words in the same list began with the same letter. (*c*) In each list, the words were distributed approximately equally among four frequency-of-occurrence categories. These four categories were: words occurring more than 100 times per million, those occurring 14 to 16 times per million, those occurring 4 times per million, and those occurring 4 times per 18 million. Apart from these restrictions, the words were assigned to different lists in a haphazard fashion.

TABLE 1

LISTS OF WORDS USED IN THE EXPERIMENT

1	2	3	4
accent	amice	action	answer
barrack	bridle	bandage	buyer
center	canard	country	cherub
drumlin	daddy	dipper	despot
entry	express	effort	ether
finding	flower	farrow	fasces
garden	gable	gambler	gorget
hoyden	hormone	hamlet	hermit
issue	impact	island	journal
jungle	juror	kitchen	letter
kernel	kitty	legion	mantel
lagoon	lactose	miser	natron
maxim	midden	noodle	orphan
newell	novice	octroi	person
office	ocean	pilgrim	question
pomade	quarter	quinsy	rennin
quillet	rumor	rennet	satin
relique	trochee	stamen	tempest
surtout	union	trollop	umbra
treason	virgin	voter	vulture
valley	wafer	waiver	windrow
walker	zither	zenith	xylem

The order of words within a given list varied systematically from trial to trial. Twenty-two different orders were constructed according to the method used in a previous paper (Tulving, 1962a). Each word appeared in each serial position just once over the total block of 22 trials, and was preceded and followed immediately, as well as by lags of 1, 2, 3, etc., by every other word in the list just once. This arrangement provides for complete absence of second-order sequential constraint among items in input lists over the block of 22 trials, and it minimizes all higher-order redundancies. The sequence in which different input orders of a given list were presented to subjects on successive trials was also systematically varied. Every subject received the input orders on successive trials in a different sequence.

The sizes of each of the four subsets defined by the TTT analysis were determined for all 22 pairs of trials (Trials 0 and 1, 1 and 2, 2 and 3, . . . 21 and 22), separately for each subject. The mean sizes of the four subsets, based on data from all 32 subjects, were then calculated. In Figure 2, these mean sizes of the four subsets are plotted against trials, yielding four different curves.

The intertrial retention (*CC*) curve is the only one that rises systematically over trials. It is a monotonically increasing negatively accelerated curve. The intratrial retention (*NC*) curve is an approximately linear curve with a negative slope. The intertrial forgetting (*CN*) curve is obviously nonmonotonic: it rises for the first few trials and then, apart from the presumably "chance" deflection upward at Trial 8, remains practically parallel with the abscissa. The intratrial forgetting (NN) curve is a monotonically decreasing curve with a negatively accelerated slope.

Both the theoretical and actual values for CC and CN on Trial 1 are 0, since subjects can neither retain nor forget any items from Trial 0 to Trial 1. These values are not shown in Figure 2. The value of NC on Trial 1, on the other hand, is determinable, and so is the value of NN. N_0C_1 corresponds to the performance score on Trial 1: $N_0C_1 = P_1$ and $N_0N_1 = L - N_0C_1$.

On inspection of the data it appeared that, as a first approximation, the intertrial retention (CC) curve could be described as a linear function in log n. The relation appears in Figure 3. The ordinate shows the mean number of items in the CC category in the recall of 32 subjects, the abscissa refers to trials on a logarithmic scale.

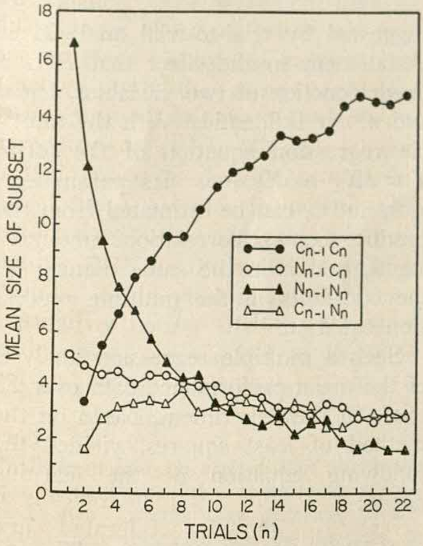


FIG. 2. Mean sizes of subsets corresponding to four different components of performance derived from trial-to-trial analysis of recall data.

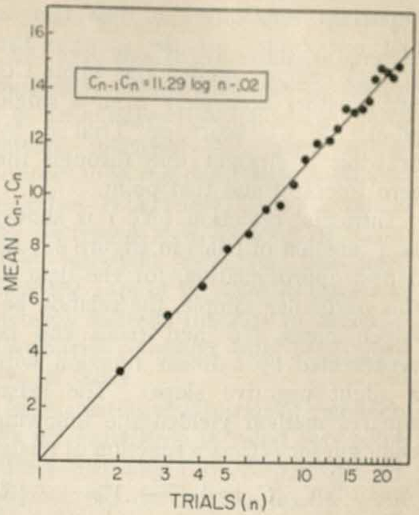


FIG. 3. Intertrial retention as a function of trials. (Trials on the abscissa are presented on the logarithmic scale. The ordinate represents the mean number of words recalled on two consecutive trials, $n-1$ and n. The straight line was fitted to the data by the method of least squares.)

The least-squares method yielded the following equation for CC as a function of trials:[5]

$$C_{n-1}C_n = 11.29 \log n - .02 \quad [2]$$

This equation is represented by the straight line drawn in Figure 3. The fit is obviously very good. It is interesting to note that the curve intersects the ordinate corresponding to $n = 1$ at a CC value of 0 as it should, since, by definition, C_0C_1 must be zero. An important implication of this fact is that one could, for the data in the present sample, predict the whole *CC* curve rather accurately from a single point, the *CC* score on Trial 2, by drawing a straight line through the zero intersect and that point.

---

[5] All equations fitted to the data can be considered to hold only within the limits of the number of trials used in the experiments. The CC function, for instance, cannot rise indefinitely, since the length of the list will impose a limit on the subjects' recall. This limit, of course, does not characterize the subjects' information-processing ability but only the experimental conditions under which this ability is permitted to operate (cf. Waugh, 1962a).

Intratrial retention (*NC*) is shown as a function of trials in Figure 4. As a first approximation, for the data in this particular sample, the relation between mean *NC* and trials can be represented by a linear function with a slight negative slope. The least-squares method yielded the following equation for *NC* as a function of trials:

$$N_{n-1}C_n = 4.87 - .10n \qquad [3]$$

The next step was to add alegbraically, on the basis of the definition of performance given in Equation 1, the intertrial and intratrial retention functions to obtain the equation of the trial-by-trial learning curve. The expression on the right-hand side in Equation 4 is the algebraic sum of corresponding expressions in Equations 2 and 3:

$$P = 11.29 \log n - .10n + 4.85 \qquad [4]$$

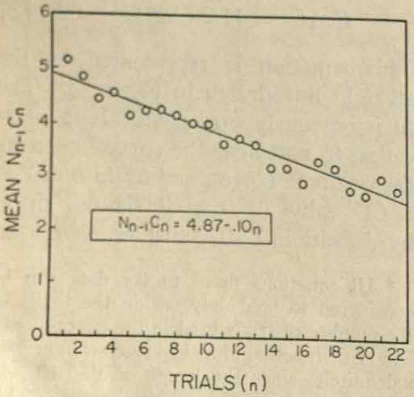


FIG. 4. Intratrial retention as a function of trials. (The ordinate represents mean number of words recalled on Trial *n* but not Trial $n-1$. The straight line was fitted to the data by the method of least squares.)

We have now "synthesized" the learning curve from two empirically determined components of performance, intertrial and intratrial retention. How well does such a curve fit the performance data? Equation 4 is graphically depicted in Figure 5, together with mean performance data on 22 trials. The fit appears good. The standard error of estimate, $\sigma_{P.n}$ was found to be .246.

Let us now turn to the problem of the mathematical analysis of the learning curve into its two component functions, without going through the relatively tedious operation of counting items in the *CC* and *NC* categories in the subjects' recall protocols. The mathematical analysis is, in a sense, a complementary operation to the synthesis of the learning curve from the empirically determined *CC* and *NC* functions.

The logarithmic function of the general form,

$$P = a \log n + bn + c \qquad [5]$$

suggested by trial-to-trial analysis, is a statement to the effect that *P* is a linear function of two variables, log *n* and *n*. It is identical with the multiple regression equation of the form, $Y = aX_1 + bX_2 + c$. Its parameters, *a*, *b*, and *c*, can be estimated from the product-moment correlations among *P*, log *n*, and *n* in the same manner as the constants in the multiple regression equation.

Such a multiple regression analysis of the mean performance data over 22 trials in the experiment, based on the method of least squares, yielded the following equation of the learning curve:

$$P = 10.70 \log n - .073n + 5.07 \qquad [6]$$

Equation 6 is to be compared with Equation 4, in which the parameters

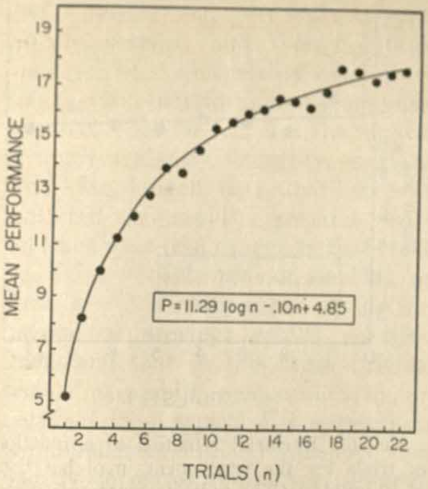


FIG. 5. The logarithmic learning curve. (The smooth curve represents the function obtained by the algebraic addition of the intertrial and intratrial retention functions shown in Figures 3 and 4. The data points represent mean number of words recalled by 32 subjects.)

were estimated from the empirically determined *CC* and *NC* components of performance. The agreement between the two sets of values, while not perfect, is close enough to suggest that they represent the same population parameters and that the differences are attributable to errors of measurement. In fact, it can be shown that with infinitely large samples in which sampling error is eliminated, the two procedures, TTT analysis and regression analysis, would yield identical results, provided of course that the forms of the functions correspond to those in the present sample, linear in log *n* for *CC* and linear in *n* for *NC*.[6]

INTRATRIAL RETENTION AND PRACTICE

We have seen that both logically and empirically performance can be shown to be an additive function of two components, intratrial and intertrial retention. For the data from our demonstration experiment, intratrial retention decreased as a function of trials, while intertrial retention increased. Since the rate of increase in intertrial retention exceeds the rate of decrease in intratrial retention, total recall goes up over trials.

In evaluating the intratrial retention data we must keep in mind a possible artifact that might be responsible for the observed decrements in *NC*. As the subject's intertrial retention increases, there remain fewer and fewer items in the list that the subject could retain from the input phase of a given trial. This would be particularly true for short lists that are mastered relatively quickly. It is possible, therefore, that the negative slope of the mean *NC* function, whenever it occurs, is attributable to the combined effects of limited list length and rapid learning by some subjects, and that it does not reflect the relation between practice and intratrial retention faithfully.

Three observations bearing on this issue should be mentioned. The first concerns the *NC* functions for "fast" and "slow" learners in the experiment described above. The equation of the *NC* function for the subgroup of 16 fast learners (those with the mean *P* scores above the group median) turned out to be $N_{n-1}C_n = 5.4 - .18n$, while for the 16 slow learners it was $N_{n-1}C_n = 4.4 - .03n$. Thus for those subjects whose performance is presumably less affected by the list length, the *NC* function is almost flat.

---

[6] While it can be argued that the main value of equations fitted to learning curves lies in new insights they generate, rather than in their ability to describe the data, it is not without interest that the logarithmic curves do seem to fit the FR data quite well. The multiple regression method was used to generate learning curves, of the form given by Equation 5, for five other sets of FR data collected at Toronto. The goodness of fit of these curves was then compared with that provided by exponential functions of the form, $P = a - be^{-cn}$ (Murdock, 1960). In each case, the standard error of estimate was much lower for the logarithmic than the exponential curve, the difference being primarily attributable to the superior ability of the logarithmic curve to describe data from the first few trials.

The second source of evidence supporting the hypothesis that the negative slope of the *NC* function represents an artifact lies in the findings of an unpublished experiment conducted at Toronto. In this experiment, 24 subjects learned three successive lists of 52 words each by the method of FR. The list words covered a wide range of Thorndike-Lorge (1944) frequencies. They were presented to the subjects on eight trials, at the rate of one second per word. The subjects had 104 seconds at the end of each input phase to recall as many words as they could. Their oral recall was recorded on tape and later transcribed.

In this experiment, no subject ever recalled more than 37 items correctly on any single trial, and mean *P* on the last trial, Trial 8, was only 26.2. The *NC* data from this experiment, even when averaged over all subjects, should thus be relatively free from any artifactual constraints. These *NC* data, pooled for the 24 subjects over all three lists, are shown in Figure 6.

The slope of the function is still negative, but it is smaller than that in Figure 4, namely −.05. More important, the negative slope is largely produced by the mean *NC* value on the very first trial. There seems to be little systematic change in *NC* over Trials 2 to 8. The finding that *NC* on Trial 1 is higher than on all subsequent trials has occurred in all samples of data that we have examined to date, and eventually we will have to come to grips with the theoretical problems that this finding entails. For the time being, however, we can conclude that for long lists intratrial retention seems to be essentially constant over trials.

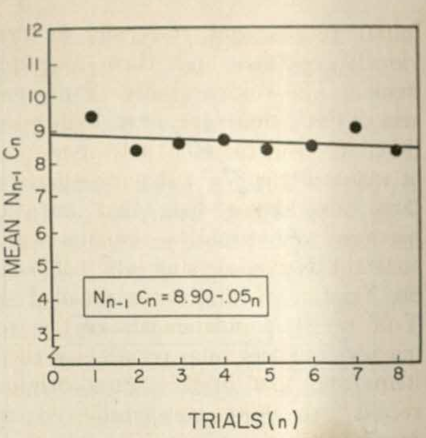


FIG. 6. Intratrial retention as a function of trials for the experiment involving lists of 52 words. (The ordinate represents mean number of words recalled on Trial *n* but not on Trial $n-1$. The straight line was fitted to the data by the method of least squares.)

The third observation relevant to the hypothesis comes from an interesting experiment reported by Murdock and Babick (1961). These investigators presented a "critical word" (CW) to the subjects on a number of successive trials, but always in the context of different lists. They found that the probability of recall of CW was not influenced by the frequency of previous presentations. This probability can be regarded as a measure of intratrial retention, estimated under conditions where list length does not impose any artifactual limits on recall. Since it did not change as a function of trials, we can conclude that intratrial retention is independent of practice.

It must be noted that Murdock and Babick examined only the probability of initial recall of their critical words while our *NC* measure includes both initial recalls and re-recalls of previously recalled and then forgotten items. The comparability of the two sets of data, therefore, may seem questionable. But the fact that the number of items in the *NC* category—particularly for longer lists that are not mastered very rapidly—remains nearly constant over trials suggests that recall on Trial *n* of items not recalled on Trial $n - 1$ is independent of the frequency of previous recalls of these items and that in this sense original recalls are quite comparable to re-recalls. In a typical FR experiment the number of initial recalls per trial usually decreases as practice proceeds, and the number of re-recalls increases. The total number of all items in the *NC* category, however, remains practically invariant at all stages of practice.

At first glance the conclusion that recalls of items on Trial *n* following nonrecall on Trial $n - 1$ is independent of frequency of previous recalls may appear at variance with the well-known fact that the probability of recall of an item is a direct function of the number of times that it has been recalled before (e.g., Miller & McGill, 1952; Underwood, 1954). Furthermore, Waugh and Smith (1962) have shown that the probability of recall of an item on Trial *n* following its nonrecall on Trial $n - 1$ does depend on the frequency of previous recalls.

The contradiction, however, may be more apparent than real. The probability of occurrence of events of a given class may increase, but if, at the same time, the size of that class decreases, the frequency of occurrence of these events may not change at all, or even change in the opposite direction. For instance, an item that has been recalled, say, on eight successive trials, and then forgotten on the ninth trial, may be recalled again on the tenth trial with a very high probability, but if recall on the ninth trial depends on the frequency of recall on the previous eight trials, there will be very few items that are forgotten on the ninth trial. In short, the apparent contradiction between Waugh and Smith's conclusion and the hypothesis that intratrial retention is independent of frequency of previous recalls remains a contradiction only as long as we confuse relative frequency, that is probability, with absolute frequency of occurrence.

In the light of these observations, then, it seems reasonable enough to entertain the hypothesis that in an ideal FR experiment, in which the subject's performance is determined only by the capacity of his memory, intratrial retention is independent of practice. The possibility, mentioned earlier, that the *NC* component underestimates intratrial retention and that intratrial retention may in fact increase over trials, however, cannot yet be completely ruled out.

INTERTRIAL RETENTION AND SUBJECTIVE ORGANIZATION

There is relatively little that needs to be said about intratrial retention in this paper. We have argued, in the light of empirical data, that intratrial retention remains essentially invariant over trials, and invariances of nature do not require explanations in the same sense as do variances. There are, however, several implications that the alleged invariance in intratrial retention has for the theory of FR learning. One such is provided by the fact that if intratrial retention is independent of trials, it must also be independent of intertrial retention. Regardless of the number of items that the average subject remembers from his previous recall, he always remembers the same number of "new" items from the immediately preceding input, where new items are those that did not occur in the immediately preceding output. Recall of the "old" does not seem to interfere with recall of the new. The exact significance of this and other implications may become clearer when we understand the processes involved in intratrial and intertrial retention. For the present we can only conclude that the limitation found in immediate memory (Miller's magical number seven, 1956b) also seems to apply to memory for new items on later trials.

In this last section of the paper we return to the problem of improvement in recall over trials. In the light of the preceding analysis this has now become the problem of increments in intertrial retention over trials. Even though the TTT analysis has contributed to the conceptual clarification of the nature of the traditional learning curve, it cannot illuminate the processes responsible for various components of recall performance. Some insight into the process of intertrial retention, however, can be derived from certain additional findings from the demonstration experiment described earlier in this paper. These findings pertain to subjective organization.

Subjects do not only recall more and more items from the input list as practice proceeds in the FR experiment. They also impose an increasingly tighter sequential organization on the recalled material. It is this organization, absent in input lists and present in output lists, that is referred to as subjective organization (*SO*). The concept has been discussed and a method for its measurement has been described in a previous paper (Tulving, 1962a). In that paper it was also argued that trial-by-trial increments in performance are functionally dependent upon organization.

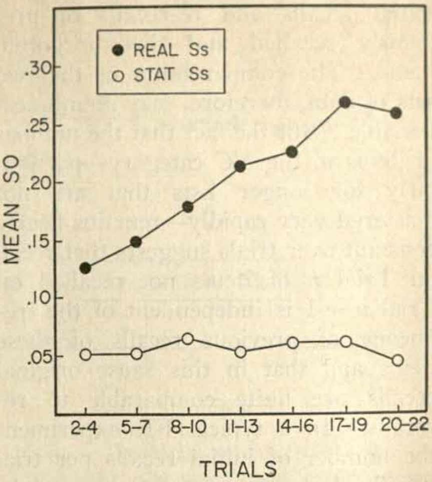


FIG. 7. Mean subjective organization (*SO*) scores on successive blocks of three trials for 32 experimental and 32 statistical subjects.

This argument can now be further evaluated by examining the relations between *SO* on the one hand and the components of performance that have emerged from the TTT analysis, intratrial and intertrial retention, on the other hand. In the demonstration experiment, *SO* (Lag 0) was calculated for each of 32 subjects on successive blocks of three trials (Trials 2–4, 5–7, . . . 20–22), as well as on blocks of seven trials (Trials 2–8, 9–15, 16–22).

Figure 7 shows the mean *SO* (Blocks = 3) for seven successive blocks of trials. It also shows mean *SO* scores for 32 statistical subjects, whose performance was matched with that of the experimental subjects, subject by subject and trial by trial. It can be readily seen that *SO* increases lawfully over trials for the real subjects, but not for the statistical subjects.

Figure 8 shows two parallel effects

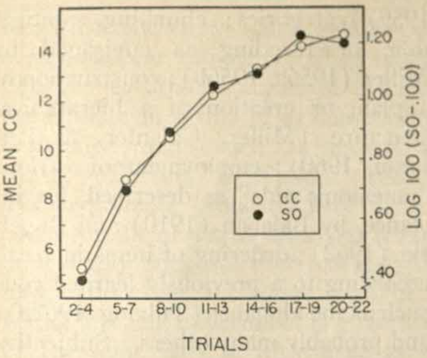


FIG. 8. Mean intertrial retention and a logarithmic transformation of mean subjective organization (*SO*) scores for blocks of three trials. (The left hand ordinate represents intertrial retention scores, the right hand ordinate shows the values of log 100 [mean $SO - .100$].)

of practice, increasing intertrial retention and increasing *SO*. Mean *CC* scores are plotted against blocks of three trials, together with a logarithmic transformation of the mean *SO* scores. The correspondence between the two curves is very good, the product-moment correlation for the seven pairs of mean scores being +.996. The important fact, of course, is that both variables increase monotonically over trials; the close correspondence between the two curves in Figure 8 depends on the transformation of *SO* scores and on the appropriate placing of the values on the two ordinates.

Does this high degree of correlation between *SO* and *CC* still occur when trials are held constant and the correlation coefficient is computed for subjects? Two variables that covary for experimental conditions need not covary for subjects (Mandler, 1959). To examine the interrelations among different response variables available from the experiment, product-moment correlations of various measures were calculated for all 32 subjects. For each subject, mean *P*, mean *CC*, mean *NC*, and log *SO* (Lag 0) were calculated for three successive blocks of seven trials, omitting Trial 1. These data are summarized in Table 2.

In interpreting these correlation coefficients, we must remember that as subjects approach the limit of performance determined by the list length L, their *CC* scores must necessarily become higher and *NC* scores lower. Thus, the negative correlations between *CC* and *NC*, as well as between *NC* and other measures, on Trials 9–15 and 16–22, are probably to a large extent artifactual. Data from the first block of trials may be less biased in reflecting the relations among these response variables. It should also be

TABLE 2

PRODUCT-MOMENT CORRELATION COEFFICIENTS AMONG MEAN *P*, MEAN *CC*, MEAN *NC*, AND LOG *SO* (LAG *O*) ON THREE BLOCKS OF SEVEN TRIALS FOR 32 SUBJECTS

Variables	Blocks of trials		
	Trials 2–8	Trials 9–15	Trials 16–22
Log *SO* and mean *P*	+.506	+.722	+.843
Log *SO* and mean *CC*	+.584	+.769	+.862
Log *SO* and mean *NC*	−.075	−.783	−.839
Mean *CC* and mean *NC*	−.018	−.881	−.943
Mean *CC* and mean *P*	+.948	+.986	+.988
Mean *NC* and mean *P*	+.295	−.792	−.882

remembered that correlations between *P* on the one hand and *CC* and *NC* on the other hand are, in a sense, spurious correlations, since *CC* and *NC* are subscores of *P*. These correlations, as well as some others, are included to complete the table. Since $P = CC + NC$, multiple correlations among *P*, *CC*, and *NC* are necessarily unity.

Considering the initial block of trials only, we note first that the correlation between *CC* and *NC* is essentially zero. Subjects' "ability" of retaining items from one trial to the next does not seem to be related to their ability of retaining items within a trial.

The second observation of interest in Table 2 concerns the positive and significant correlation between *CC* and log *SO*, and the absence of any notable correlation between *NC* and log *SO*, in the initial block of seven trials. For any given subject, therefore, intratrial retention is independent of the amount of organization that the subject imposes on the material, while intertrial retention seems to be directly related to *SO*.

At the empirical level, *SO* refers to the subjects' tendency to recall certain items in close temporal contiguity to one another. At the conceptual level, this tendency can be thought to represent the formation and existence of higher-order memory units. It is as if the list items—all already in the memory storage prior to the experiment, as was argued earlier—are rearranged in the storage in the course of trial-by-trial practice. Such rearrangement manifests itself and can be described in a variety of ways—development of associations of the type that define the associative meaning of a word (Deese, 1962); clustering in terms of conceptual (Bousfield, 1953; Cohen, 1963), associative (Jenkins & Russell, 1952), or synonymic (Cofer, 1959) categories; chunking, unitization, or recoding as envisaged by Miller (1956a, 1956b); construction of a plan, or creation of a hierarchical structure (Miller, Galanter, & Pribram, 1960); employment of various "mnemonic aids" as described, for instance, by Balaban (1910) and Bugelski (1962); ordering of items in recall according to a previously learned code such as the alphabet (Tulving, 1962b); and probably many others. Subjective organization is just a general name for all of these processes. To the extent that the higher-order memory units that result from organization are not specified experimentally and to the extent that the nature of the units is determined by the subject's previous experience, it might be justifiable to refer to these units as subjective units (S units). The functional significance of the development of these units lies in the increased accessibility of individual items constituting a unit. An item can be retrieved on its own merits, or through other items in its higher-order S unit.

If we accept the assumption that the measure of *SO* reflects the extent to which initially "unrelated" list items are formed into higher-order S units, it becomes possible to regard the increasing intertrial retention as reflecting nothing more or less than the increasing size of S units, and to entertain the hypothesis that the number of S units mediated by intertrial retention remains essentially constant over trials. This hypothesis agrees well with Miller's (1956a, 1956b) conception of memory system that is limited by the number of units that can be retrieved in succession without intervening external instructions, but not by the information content of these units. There is already some evidence that large differences in first-trial free recall become very much smaller or even disappear completely when retention is measured in terms of S units rather than in terms of grammatically designated units such as individual words (Tulving & Patkau, 1962). The hypothesis proposed here to account for the observed intertrial retention simply suggests that handling of information over successive trials in a verbal learning experiment is governed by the same laws of invariance as is handling of information within a single trial.

SUMMARY

1. A subject attempting to memorize a list of previously integrated items under the conditions of a multitrial free recall experiment "learns" every single item at the time of its presentation. Because of rapid intratrial forgetting only some items can be recalled in the output phase of the first trial. The number of such items represents a measure of intratrial retention.

2. Recall on the second trial, as well as on all subsequent trials, can be analyzed into two components—intratrial and intertrial retention. Intratrial retention refers to retention of items from the input phase of the trial, intertrial retention refers to retention of items from the output phase of the previous trial.

3. The traditional learning curve can be expressed as an additive function of intratrial and intertrial retention curves plotted against trials. For many samples of data, one of which was considered in detail in the paper, intertrial retention increases as a logarithmic function of trials, whereas intratrial retention decreases as a linear function of trials. The logarithmic learning curve of the form, $P = a \log n - bn + c$, not only describes the recall performance quite adequately, but also makes explicit the two components of performance, $a \log n$ representing intertrial retention and $c - bn$ representing intratrial retention.

4. Although the intratrial retention component of performance has been observed to decrease as a function of trials, it can be argued, on the basis of several kinds of evidence, that this decrease is an artifact attributable to limited list length used in most experiments, and that in an unconstrained task involving memorization of a list of words, intratrial retention is invariant at all stages of practice.

5. Supported by the finding that intertrial retention is positively correlated with the amount of organization that subjects impose on their recall, the hypothesis was advanced that the increase in intertrial retention reflects the increase in the size, but not the number, of the higher-order subjective units of material which can be retrieved from storage.

REFERENCES

BALABAN, A. Über den Unterschied des logischen und mechanischen Gedächtnisses. *Z. Psychol.*, 1910, 56, 356–377.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, 49, 229–240.

BROADBENT, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

BROWN, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, 10, 12–21.

BRUNER, J. S., MILLER, G. A., & ZIMMERMAN, CLAIRE. Discriminative skill and discriminative matching in perceptual recognition. *J. exp. Psychol.*, 1955, 49, 187–192.

BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *J. exp. Psychol.*, 1962, 63, 409–412.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

Cofer, C. N. A study of clustering in free recall based on synonyms. *J. gen. Psychol.*, 1959, **60**, 3–10.

Cohen, B. H. An investigation of recoding in free recall. *J. exp. Psychol.*, 1963, **65**, 368–376.

Conrad, R., & Hille, Barbara A. The decay theory of immediate memory and paced recall. *Canad. J. Psychol.*, 1958, **12**, 1–6.

Deese, J. On the structure of associative meaning. *Psychol. Rev.*, 1962, **69**, 161–175.

Ebbinghaus, H. *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie.* Leipzig: Duncker & Humbolt, 1885.

Estes, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

Estes, W. K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika*, 1961, **26**, 73–84.

Feigenbaum, E. A., & Simon, H. A. A theory of the serial position effect. *Brit. J. Psychol.*, 1962, **53**, 307–320.

Horowitz, L. M. Free recall and ordering of trigrams. *J. exp. Psychol.*, 1961, **62**, 51–57.

Jenkins, J. J., & Russell, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.*, 1952, **47**, 818–821.

Mandler, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

Mandler, G. Stimulus variables and subject variables: A caution. *Psychol. Rev.*, 1959, **66**, 145–149.

McGeoch, J. A. *The psychology of human learning.* New York: Longmans, 1942.

Miller, G. A. Human memory and the storage of information. *IRE Trans. Inform. Theory*, 1956, **2**, 129–137. (a)

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–96. (b)

Miller, G. A., Galanter, E., & Pribram, K. *Plans and the structure of behavior.* New York: Holt, 1960.

Miller, G. A., & McGill, W. J. A statistical description of verbal learning. *Psychometrika*, 1952, **17**, 369–396.

Murdock, B. B., Jr. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–234.

Murdock, B. B., Jr. Short-term retention of single paired associates. *Psychol. Rep.*, 1961, **8**, 280. (a)

Murdock, B. B., Jr. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618–625. (b)

Murdock, B. B., Jr. Short-term retention of single paired associates. *J. exp. Psychol.*, 1963, **65**, 433–443.

Murdock, B. B., Jr., & Babick, A. J. The effect of repetition on the retention of individual words. *Amer. J. Psychol.*, 1961, **74**, 596–601.

Peterson, L. R., & Peterson, Margaret J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

Peterson, L. R., Peterson, Margaret J., & Miller, A. Short-term retention and meaningfulness. *Canad. J. Psychol.*, 1961, **15**, 143–147.

Peterson, L. R., Saltzman, Dorothy, Hillner, K., & Land, Vera. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396–403.

Postman, L. One-trial learning. In C. N. Cofer & Barbara S. Musgrave (Eds.), *Verbal behavior and verbal learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 295–321.

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.*, 1962, **69**, 344–354. (a)

Tulving, E. The effect of alphabetical subjective organization on memorizing unrelated words. *Canad. J. Psychol.*, 1962, **16**, 185–191. (b)

Tulving, E., & Arbuckle, Tannis Y. Sources of intra-trial interference in immediate recall of paired associates. *J. verbal Learn. verbal Behav.*, 1963, **1**, 321–334.

Tulving, E., & Patkau, Jeannette E. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canad. J. Psychol.*, 1962, **16**, 83–92.

Underwood, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, **51**, 276–282.

Underwood, B. J., & Keppel, G. One-trial learning? *J. verbal Learn. verbal Behav.*, 1962, **1**, 1–13.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

WAUGH, NANCY C. Free versus serial recall. *J. exp. Psychol.,* 1961, 62, 496–502.

WAUGH, NANCY C. Length of series and the learning curve. *Amer. J. Psychol.,* 1962, 75, 177–192. (a)

WAUGH, NANCY C. The effect of intra-list repetition on free recall. *J. verbal Learn. verbal Behav.,* 1962, 1, 95–99. (b)

WAUGH, NANCY C., & SMITH, J. E. K. A stochastic model for free recall. *Psychometrika,* 1962, 27, 141–154.

(Received September 14, 1962)

Psychological Review
1964, Vol. 71, No. 3, 238-242

THEORETICAL NOTES

A NOTE ON THE INCONCLUSIVENESS OF ACCEPTING THE NULL HYPOTHESIS

WARNER R. WILSON AND HOWARD MILLER[1]

University of Hawaii

When an experimenter predicts exactly the differences to be expected, his choice of analysis determines whether acceptance or rejection of a null hypothesis will lend support to his theory. He makes a similar choice when he decides to study a situation in which his theory predicts no differences vs. one in which some differences are predicted. It is better to make both choices so that support is based on rejection rather than acceptance of a null hypothesis because, in both cases, a rejection-support strategy enables one to minimize and quantify the danger of accepting an erroneous theory even in imprecise experiments.

Grant (1962) and Binder (1963) have recently discussed the utility of basing support for a theory on the acceptance of a null hypothesis (Ho) versus basing support on the rejection of a Ho. It is convenient to refer to the procedure where acceptance of a Ho leads to support for a theory as acceptance-support (a-s), and to the procedure where support for a theory comes from rejection of a Ho as rejection support (r-s). Both writers have contributed to the clarification of the issue, but ambiguity and disagreement still remain.

Before discussing the problem it may be desirable to discuss some of the terms to be used. This paper follows the lead of Grant (1962) in viewing the Ho as a supposition to the effect that no difference or no relationship exists or that no effect is exerted. Seemingly the term null hypothesis is often used in this way, and certainly Grant (1962) used it in this way when he first stated the issue. It is therefore, in this sense, a statistical hypothesis to whose rejection alone a probability estimate may be placed. As Binder points out, those of the Yule-Fisher school view the investigator as having two alternatives: he either rejects the Ho or reserves judgment. Those following the lead of Neyman and Pearson either reject the Ho or accept it.[2] The Yule-Fisher school would seem more compatible with a-s since in a-s one either rejects the Ho or withholds judgment. Likewise, the Neyman-Pearson school seems more compatible with a-s since it allows acceptance of the Ho, and, in a-s it is the acceptance of the Ho that provides support for a theory which predicted no difference in the first place. However, any reasonable person would concede that if the theory predicts no difference, the failure to find a significant difference gives some comfort to the theory, at least relatively speaking. It would seem, therefore, that it makes little difference which of these views one takes.

This paper suggests that the issue of a-s versus r-s strategies is broader than Grant and Binder indicate. It is hoped that the additional perspectives presented here will add further clarification and help provide grounds for the resolution of the issue. Grant and Binder discuss primarily a strategy decision which must be made in relation to the analysis of data. This paper points out that a similar decision must be made in relation to the collection of data.

In some cases, experimenters employ theoretical models to predict expected outcomes exactly. Grant and Binder direct most of their attention to situations of this kind. They point out that when evaluating the correspondence between predictions and outcomes, the experimenter has at least two choices: he can test the Ho of no difference between predicted and actual values, or he can test the Ho of no significant correspondence between observed and predicted values. In the first case, support for the theory comes from acceptance of a Ho; in the second case, support comes from rejection of a Ho. It may be noted that in this context the experimenter commits himself to an a-s or an r-s strategy when he decides how to analyze his data. It may be convenient to refer to the two types of analyses as acceptance-support-analysis (a-s-a) and rejection-support-analysis (r-s-a). A typical a-s-a would involve showing, perhaps by means of *t*-tests, that a number of predicted points do not deviate significantly from the points predicted by a theory and concluding that the theory is, therefore, supported. An r-s-a of the same data would involve showing a significant correspondence between predicted and observed points by means of correlation or analysis of variance as discussed by Grant.

---

[1] Now enrolled in the School of Medicine at Western Reserve University.

[2] Binder's (1963) article provides a more detailed and well-documented discussion of these two viewpoints.

Grant (1962) makes several objections to a-s-a: It is seldom, in psychology at least, that an experimenter can have real faith in the accuracy of his theory. Hence, if an experimenter finds no difference between theoretical and empirical values, that is, if a-s is obtained, this result usually merely indicates that the experiment was not sufficiently sensitive to indicate the imperfections in the theory that are almost certainly there. Further, the more insensitive the experiment, the more likely it is to "support" a theory even if the theory is poor. When one seeks r-s, however, insensitivity works against acceptance of a theory. Furthermore, it is also difficult to give a satisfactory interpretation to a failure to obtain a-s because, if an experiment is very sensitive, it may deny a-s to a theory which is really fairly good. When an experimenter seeks r-s, however, sensitivity increases his chances of obtaining support.

Binder (1963, cf. p. 111) claims that parallel objections can be made about r-s-a. An r-s-a may fail to support a good theory if the experiment is very insensitive; and, at the same time, if the experiment is very sensitive, an r-s-a may lend support to a worthless theory, that is, one that has negligible predictive utility. Grant's arguments, says Binder, do not add up to an objection against an a-s strategy; they add up to an objection against experiments that are either too insensitive or too sensitive. Experiments whose sensitivity is appropriate to the present stage of theory construction and application will enable an experimenter to accept helpful theories and reject useless ones no matter which type of analysis he applies.

While this last point may be sound enough logically, it does not help much practically, since it would seem to be rather difficult to establish any workable formula for deciding what is too precise or imprecise in a given setting. Better to have a strategy that protects against the more serious type of error even if optional sensitivity is not maintained.

Binder (1963) does make a contribution, however, in that he brings to our attention that we can make several different kinds of mistakes. We can (*a*) accept a false theory; (*b*) withhold judgment about or reject a good theory; or (*c*) accept a "true but poor" theory (as discussed by Binder, 1963, pp. 111–112). It may be helpful to note the relation among choice of a-s-a or r-s-a, use of sensitive or insensitive experiments, and the type of error one is subject to:

In an imprecise experiment a-s-a leads to Error Type *a*

In a precise experiment a-s-a leads to Error Type *b*

In an imprecise experiment r-s-a leads to Error Type *b*

In a precise experiment r-s-a leads to Error Type *c*

The choice between a-s-a and r-s-a would seem to hinge on an evaluation of the seriousness of these different mistakes.

Traditionally, Type *a* is considered much more serious than Type *b*. Campbell (1959) has presented a very sophisticated justification of this time-honored assumption. Ordinarily experimenters pay no attention to Type *c;* most experimenters are content to base r-s on a significant *t* or *F* and to worry not at all if the effect is so small as to be negligible. It is apparently quite rare in psychology for an experimenter to insist that a difference be larger than some minimum before it is accepted as important. And, many psychological theories are able to predict only that differences, of unspecified magnitude, should or should not occur. Some even argue that small effects may be very important from the viewpoint of theory development and maintain, in effect, that when one reports a significant but very small effect he is not making an error at all but being highly virtuous. Nonetheless, it may be unwise to give so little heed to the size of effects, that is, to Error Type *c*. It could be argued that psychologists may be spending a great deal of their time studying variables better viewed as irrelevant. This would seem to be more a fault of *interpretation* of a detected relationship than its detection itself. Finding a relationship is still a virtue while misusing it may not be.

One could even argue that Error Type *a,* acceptance of a false theory, is much the same as Error Type *c,* acceptance of a theory that is true as far as it goes but which is able to predict only a negligible part of the variance in the data. Indeed, Binder presents much of his defense of a-s on the fact that while a-s may lead to *a,* r-s has a corresponding weakness in that it may lead to *c*. Even if it is granted that Error Types *a* and *c* are similar in their consequences—and this is a concession that psychology as a whole apparently does *not* make—one can still note grounds for preferring precise experiments and an r-s analysis to imprecise experiments and a-s analysis. For one thing, human nature being what it is, more experimenters will probably be tempted to do imprecise experiments than precise experiments. One might argue, however, that this consideration is beside the point. More important is the fact that the precise experiment necessarily provides data to indicate how adequate the theory in question is. To be sure, it is a peculiarity of psychological data reporting that significance of relationship is always indicated while degree of relationship usually is not. Nonetheless, if the degree of relationship is slight, this can readily be inferred if the appropriate computations, or even inspections, are made. The experimenter may accept a theory that predicts very little but he is not likely to be seriously deluded about how much it predicts. One is much more in the dark with the imprecise a-s-a, for in this case the data show that the theory may be perfect, but, in the absence of r-s-a it may, as far as anyone can tell, also be very poor or even completely worthless.

These various considerations would seem to support Grant's contention that the usual acceptance-rejection tactics might well be replaced or at least supplemented by a quality control procedure which would indicate not merely if theories predict but how well. The above considerations would not, however, seem to support Binder's contention that a-s and r-s expose the experimenter to completely parallel and equal serious pitfalls. Indeed, in light of the various stated reasons for viewing Error Type *a* as more serious than Error Type *c* (which may not even in itself be an error), it would seem that an evaluation of the various considerations Binder raises would lead to a preference for an r-s rather than an a-s strategy.

Perhaps some different perspectives can clarify further the basis for this preference. Suppose the same set of data provides both types of support or neither. In such cases it would make little difference which analysis was applied. But suppose the data provide r-s but not a-s. Such data can be meaningfully interpreted as lending the theory some support as Grant (1962) has pointed out. The critical question is, "What are the implications of accepting a-s when r-s cannot be claimed?" It is suggested that a-s without r-s means nothing. This result can seemingly only occur when the experiment *is* too insensitive to justify any conclusion. Although an a-s analysis may sometimes be a useful or interesting supplement to the r-s analysis, Grant (1962) seems justified in maintaining that a-s alone is not enough. To state this conclusion in another way: a-s-a is adequate in some situations but in all these cases r-s-a will be adequate also, however, in certain other situations, which apparently cannot always be anticipated, r-s-a will be adequate but a-s-a will lead one astray. Hence, why use a-s-a?

Thus far this paper has considered only cases in which the experimenter is comparing the fit between data points and theoretical expectations that take the form of exact quantitative predictions, e.g., the correspondence between an empirical and predicted learning curve. It is hopefully clear that two different approaches can be taken to the analysis of such data, the acceptance-support-analysis or the rejection-support-analysis.

Obviously many, if not most, experiments in psychology do not involve exact predictions. Often it is only the directionality of results that is predicted, e.g., the experimental group is expected to be less than, greater than, or perhaps equal to the control group. Do the issues raised by Grant and Binder in the context of exact quantitative prediction have any relevance to experiments that test theories which predict only whether groups are less than, greater than, or equal to each other? It is suggested that this second context also allows the experimenter to adopt an a-s or an r-s strategy and that the arguments favoring r-s are just as compelling in this situation.

It is true that when prediction is not exact, the kind of analysis applied will almost necessarily be analogous to what has been discussed as r-s-a. Suppose one predicts that the experimental group will exceed the control group. The Ho states no difference. Support for the prediction comes from rejecting this Ho. Application of an a-s-a, in this context, would involve conceptualizing the Ho as follows: "The groups do not differ significantly from the rank order predicted." A-s is obtained unless the groups differ significantly in the direction opposite that predicted. One, of course, is not likely to think of an a-s-a in this particular context because, in this context, a-s is ridiculously easy to obtain. Perhaps, however, this fact only lends support to the assertion made earlier—a-s without r-s means nothing.

Even when predictions are only directional and an r-s-a is used, the experimenter can still adopt an acceptance versus a rejection support strategy by choosing to design his experiment so that his theory predicts no differences versus some differences. It will be convenient to refer to the no-difference experiment as an a-s-design (r-s-d) and to the some-difference experiment as an r-s-design (r-s-d). If the experimenter chooses an a-s-d, he can claim support for his theory if no differences are found and at the same time present his negative results as evidence against any rival theories that do predict a difference. Rock and others (Rock, 1957; Rock & Heimer, 1959; Rock & Steinfeld, 1963) have used just such a strategy. If the experimenter wishes to use an r-s strategy, he adopts an r-s-d; and if differences are obtained, he claims support for his position and evidence against any theories that predict no difference. Wilson (1962) has used just such a strategy. The experimenter may, of course, change strategies in midstream. For example, if an experimenter fails to get r-s, he may adopt a new theoretical position that predicts no difference and cite his data as giving a-s to it. When Thorndike reformulated his law of effect on the basis of negative results, he essentially changed from an r-s to an a-s strategy. Seemingly, an experimenter could, in theory, start out without a theory and formulate one post hoc, in which case he might claim either a-s or r-s.

The objections made to a-s-a seem to apply with equal or greater force to a-s-d. In both cases imprecision increases the danger of accepting an erroneous theory, whereas both r-s-a and r-s-d protect against erroneous conclusions even in imprecise experiments.

A further important and compelling point would seem to be that in all of these cases we are using statistics because empirical hypotheses do not find absolute acceptance (or rejection). Rather they are accepted with more or less confidence depending on the adequacy of the evidence. Indeed, the main point of inferential statistics would seem to be to attack probability statements to the acceptance of ideas. Unless we are able to predict a necessary *degree* of correspondence or difference (which is rare in psychology), we are only able to attach such a probability statement to the *rejection* of the null hypothesis. For example, in a directional empirical prediction we can say that 1 or 5% of the time (as we choose) we will be wrong in rejecting the null hypothesis on the basis of such data as these. We cannot, however, conversely say what the probability is of falsely accepting it except for the unlikely event in which we can specify a necessary magnitude. Therefore, only an r-s strategy will supply an exact estimate and an a-s strategy allows one to "accept" his hypothesis only in the negative sense of having found no evidence against it. This does not seem to be very satisfying grounds especially when error variance is large.

A final consideration has to do with the nontheoretical value of the information obtained from an a-s-d versus an r-s-d. It is perhaps not undue to ask of an experimenter what value his information has other than its implications for choosing between theories. It may be noted that a-s-d seem to commit one to the cataloging of ineffective variables and procedures. Very likely a cataloging of effective variables and procedures, which is the logical result of r-s-d, would be more useful.

SUMMARY

A choice can be made between an acceptance-support and a rejection-support strategy when prediction is merely directional as well as when it is exact. When exact quantitative predictions are derived, one chooses an acceptance-support versus a rejection-support strategy by choosing an acceptance-support analysis versus a rejection-support analysis. When only the directionality of the outcome is predicted, a rejection-support analysis is almost sure to be used, but one can still make the choice of an acceptance-support versus a rejection-support strategy by choosing an acceptance-support design versus a rejection-support design. In both cases rejection-support seems to be the better strategy, especially in that it enables one to minimize and quantify the danger of accepting an erroneous theory even in imprecise experiments.

REFERENCES

BINDER, A. Further considerations on testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.,* 1963, **70**, 107–115.

CAMPBELL, D. T. Methodological suggestions from a comparative psychology of knowledge processes. *Inquiry,* 1959, **2**, 153–182.

GRANT, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.,* 1962, **69**, 54–61.

ROCK, I. The role of repetition in associative learning. *Amer. J. Psychol.,* 1957, **70**, 186–193.

ROCK, I., & HEIMER, W. Further evidence of one-trial associative learning. *Amer. J. Psychol.,* 1959, **72**, 1–16.

ROCK, I., & STEINFELD, G. Methodological questions in the study of one-trial learning. *Science,* 1963, **140**, 822–824.

WILSON, W. R. Learning of paired-associates: Definitely not all-or-none. *Amer. Psychologist,* 1962, **17**, 372. (Abstract)

(Received August 15, 1963)

Psychological Review
1964, Vol. 71, No. 3, 243-246

A COMPARISON OF DATA OBTAINED WITH DIFFERENT FALSE-ALARM RATES[1]

DONALD A. NORMAN

Harvard University

This note examines the comparison of performance measures when both successes (hits) and failures (false alarms) must be considered. A simple graphical procedure is developed that allows comparison to be made among experimental conditions with very limited assumptions about the underlying processes.

One recurring analytical problem is the comparison of performance measures when both successes (hits) and failures (false alarms) must be considered. An example of such a comparison occurs in the analysis of a choice experiment with, let us say, two possible responses and two stimulus conditions. If Response 1 is correct only for Stimulus A and Response 2 only for Stimulus B, then the results of the experiment are summarized by the relative frequencies of correct 1 responses—hits—$[\hat{p}(1|A)]$ and false 1 responses—false alarms—$[\hat{p}(1|B)]$. The problem is to compare the results from Experimental Condition I, $\langle\hat{p}_I(1|A), \hat{p}_I(1|B)\rangle$ with the results from a second Condition II, $\langle\hat{p}_{II}(1|A), \hat{p}_{II}(1|B)\rangle$ in order to determine which condition gives the superior performance.

This problem arises in many contexts, for example:

1. In recognition memory, the probability of recognizing old items from a list of old and new items is compared for several different experimental variables (Davis, Sutherland, & Judd, 1961; Woodworth & Schlosberg, 1954);
2. In psychophysics, the performance of an observer detecting (or identifying) a signal is compared for several different experimental variables (Green, 1960);
3. In information retrieval, several library search procedures are to be evaluated and compared when the different systems have differing correct and incorrect retrieval rates (Swets, 1963).

Early attempts to make this type of comparison assumed that errors $[\hat{p}(1|B)]$ and $[\hat{p}(2|A)]$ were either to be avoided or to be corrected (Woodworth & Schlosberg, 1954). Two types of correction formulas have been popular:

$$\text{“true rate”} = (\text{proportion correct}) - (\text{proportion incorrect}) \quad [1]$$

or

$$\text{“true rate”} = [\hat{p}(1|A) - \hat{p}(1|B)]/[1 - \hat{p}(1|B)]. \quad [2]$$

These two methods are not only different, but are also conflicting; it is quite easy to find pairs of numbers such that by using Equation 1, Condition I is "better," while by using Equation 2, Condition II is "better." These two correction techniques, the assumptions underlying them, and clear demonstrations of their incorrectness in recognition memory, audition, and vision have been studied by a number of investigators, among them Egan (1958) and Swets, Tanner, and Birdsall (1961).

Recent developments in statistical decision theory and psychophysics have provided an answer to the comparison problem when the complete operating function (ROC curve) relating $\hat{p}(1|A)$ to $\hat{p}(1|B)$ is known (see the references in Swets, Tanner, & Birdsall, 1961). Unfortunately, it is often impossible or impractical to determine this relation. In some fields of study there are mathematical models for these functions, but

---

[1] This work was performed while the author was a National Science Foundation Postdoctoral Fellow at the Harvard Center for Cognitive Studies. I would like to thank Wayne Wickelgren for his helpful suggestions.

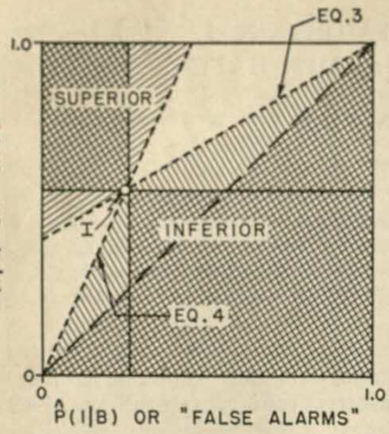


FIG. 1. Comparison of data with one point from each condition. (The shaded areas illustrate where Condition II is superior or inferior to Condition I.)

unfortunately, whenever there is one model there are usually three or four models, all different, but all equally good descriptions of the same data.

BASIC ANALYSIS

The purpose of this note is to examine some comparisons that can be made among the data with very limited theoretical assumptions. To do this, first consider what we mean by one condition being better than another. There are two requirements that seem reasonable. First is that a condition (I) is superior to another (II) if it leads to a hit rate that is better, and a false alarm rate that is lower, than that of the other condition. The second requirement is that any differences in the performance between the two conditions could not have occurred by chance alone.

Figure 1 shows the unit square representing the possible hit–false-alarm pairs. The minor diagonal (the dashed line) represents chance behavior—points below the diagonal are worse than chance behavior, and so are clearly inferior to points above the diagonal.[2] Now, let us fix the point from Condition I (hereafter referred to as I) in some arbitrary location and consider the regions in the unit square in which a second point (II) must lie in order to be unequivocally superior or inferior to Point I. By the first of our two requirements, Point II is inferior if it lies in the rectangular space below and to the right of Point I. By the same argument, if Point II lies anywhere in the rectangular space above and to the left of Point I, then II must be superior to I. These regions are illustrated in Figure 1.

Now consider the ways in which the location of the points fluctuate due to chance (or any behavior such as guessing that is independent of the conditions under study). If the number of "1" responses increases by chance, completely independent of the stimulus conditions, then Point I moves to the new Location I′ with coordinates

$$\hat{p}_{I'}(1|A)=\hat{p}_I(1|A)+t[1-\hat{p}_I(1|A)]$$

and

$$\hat{p}_{I'}(1|B)=\hat{p}_I(1|B)+t[1-\hat{p}_I(1|B)], \quad [3]$$

where $0 \leqslant t \leqslant 1$ is the proportion of the increase. Equation 3 indicates that Point I′ moves along the straight line segment connecting Point I with the coordinates $\langle 1, 1\rangle$ as t varies from zero to one. Similarly, if by chance the number of "2" responses should increase by some proportion u, completely independent of the stimulus conditions, then Point I moves to the new Location I″ with coordinates

$$\hat{p}_{I''}(1|A) = u\hat{p}_I(1|A)$$

and

$$\hat{p}_{I''}(1|B) = u\hat{p}_I(1|B), \quad [4]$$

where $0 \leqslant u \leqslant 1$. Equation 4 determines the straight line segment connecting Point I with the coordinates $\langle 0, 0\rangle$.

Now, if Point II should lie either along the lines of Equations 3 and 4 or below and to the right of these lines, it cannot be superior or equivalent to Condition I by our second requirement: the same or better result could have been obtained from Condition I by chance alone. So we may add this area to the region labeled "inferor" in Figure 1.

---

[2] Although points below the line of chance behavior are usually considered inferior to those above the line, they still represent efficient discrimination. For example, were the subject's responses always wrong, the obtained point would be at the lower right corner $\langle 1, 0\rangle$: signifying excellent—although somewhat perverse—performance.

By a similar argument, Point II is superior to I if it is located in the unit square so that Point I lies within the area enclosed by the lines connecting II with $\langle 0, 0\rangle$ and $\langle 1, 1\rangle$. It is easily seen that the region in which II will be superior to I is defined by the area above and to the left of the continuations of the line segments of Equations 3 and 4 (see Figure 1). Note that the argument is entirely equivalent to requiring that the operating characteristic be monotonic from $\langle 0, 0\rangle$ to $\langle 1, 1\rangle$ with nonincreasing slope. At any Point I, the slope of the operating characteristic is bounded by the slopes of the line connecting I with $\langle 0, 0\rangle$ (as a maximum) and the line connecting I with $\langle 1, 1\rangle$ (as a minimum). The two points cannot possibly lie on the same curve unless they both lie in the region bounded by the lines.

As we can see in Figure 1, we can make some good inferences on the relationship between two experimental conditions without any strong assumptions about the underlying processes. The unshaded areas in Figure 1 are the ambiguous regions where, in the absence of more data or theories, no conclusion can be reached.

Combining Data Points

Much stronger restrictions on the unit square can be determined if a second point is available for Condition I. With two points, I_1 and I_2, the regions where Condition II is inferior or superior to Condition I is obtained by the disjunction of the corresponding areas determined separately for each point. (If, indeed, the two points come from the same performance condition, there should be no conflicting regions.) This new area is illustrated in Figure 2 by the diagonal shading.

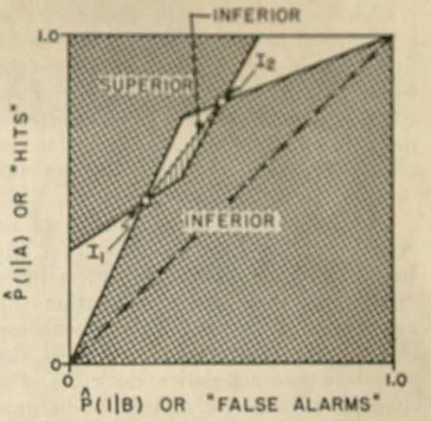


Fig. 2. Comparison of data with two points from one condition. (The shaded areas illustrate where Condition II is superior or inferior to Condition I.)

If both Points I_1 and I_2 can result from Condition I, then intermediate points can also be obtained if performance alternates between the two points. In this case, any point on the straight line segment connecting I_1 and I_2 can be obtained in Condition I. Thus, if Point II were located in the triangle bounded on the upper left by the Line I_1—I_2 it must be considered inferior to I. The addition of the second point, I_2, reduces the ambiguous region in the unit square considerably, as shown in Figure 2. Note that if two points are available for both conditions, not only does Figure 2 apply but the analytical power is strengthened. The analysis of Figure 2 should be applied twice, once for the pair of points from each condition. The results are ambiguous only if all the points from both analyses are ambiguous.

Other Considerations

The regions illustrated in the figures are derived from an exact knowledge of the points. These, of course, are seldom known; the empirically determined values are but statistical estimates of the true locations. The confidence intervals for the estimates of the two coordinates of each point determine confidence ellipses which surround the points. The confidence limits for the boundaries of the inferior and superior regions can be determined by drawing the lines of Figures 1 and 2 tangent to the outer edges of the ellipses. These confidence limits can then be used to determine the effect of statistical variability on the comparison of the data points.

One basic assumption remains to be stated. It has been assumed that the true underlying operating characteristics —whatever their shape or location—do not cross one another. This assumption is weak for two reasons. One, if the operating characteristics do cross, the concepts of superior or inferior become strained and perhaps even meaningless. Two, to my knowledge, none of the varied theories or data on operating characteristics suggest such a possibility.

CONCLUSION

The analyses in this note can easily be extended to more data points and more stimuli and responses. The important result is that experimental conditions can often be ranked with respect to performance without any assumptions about the underlying mechanisms or artificial correction techniques. Exact knowledge of the operating characteristic is always the most desirable solution and supersedes these techniques. Unfortunately, such information is not always available.

REFERENCES

DAVIS, R., SUTHERLAND, N. S., & JUDD, B. B. Information content in recognition and recall. *J. exp. Psychol.*, 1961, **61**, 422–429.

EGAN, J. P. Recognition memory and the operating characteristic. Technical Note, Contract No. AF 19(604)–1962, 15 June 1958. Bloomington: Indiana University Hearing and Communication Laboratory.

GREEN, D. M. Psychoacoustics and detection theory. *J. Acoust. Soc. Amer.*, 1960, **32**, 1189–1203.

SWETS, J. A. Information retrieval systems. *Science*, 1963, **141**, 245–250.

SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954. P. 700.

(Received November 1, 1963)

VOL. 71, NO. 4 JULY 1964

PSYCHOLOGICAL REVIEW

STIMULUS INTENSITY EFFECTS DEPEND UPON THE TYPE OF EXPERIMENTAL DESIGN[1]

G. ROBERT GRICE AND JOHN J. HUNTER

University of Illinois

In experiments dealing with variations in CS intensity in eyelid conditioning and in signal intensity in simple reaction time, it was determined that substantially greater effects are obtained if individual Ss are exposed to the different intensities than if each S experiences only one intensity value. Different results are thus to be expected from experimental designs which use these different procedures. Stimulus intensity generalization experiments are inherently of a type which should produce substantial intensity effects. The data tend to support a dynamogenic theory which does not depend upon fixed functions of stimulus energy, but in which the level of adaptation must be considered.

The intensity of the CS has usually been found to be a relatively ineffective variable in human conditioning. In eyelid conditioning, Carter (1941) and Grant and Schneider (1948) found no evidence that this variable reliably affected either response strength or the associative level of conditioning. Grant and Schneider (1949) also failed to find significant CS intensity effects in GSR conditioning. The only human data cited by Hull (1949) in his presentation of the stimulus intensity dynamism (*V*) were studies of signal intensity in reaction time (Cattell, 1886; Piéron, 1920) and the well-known GSR study of intensity generalization by Hovland (1937). Walker (1960) reported that the intensity of an auditory CS reliably affected the frequency of eyelid responses of the voluntary form, and conditioned responses of short latency only. However, the effects which she reported were rather small. More recently, Beck (1963) has obtained a result in eyelid conditioning which, in view of this earlier literature, is quite remarkable. In an investigation of possible interaction effects between CS intensity, UCS intensity, and drive level as inferred from an anxiety scale, she included the CS intensity variable in a factorial design as a within-*S*s effect. Two intensity values were used and were both administered to all *S*s in an irregular order throughout 100 conditioning trials. This was done in order to achieve the statistical precision of this design, since the effect was assumed to be rather weak. The result was that the CS intensity variable not only reached statistical significance, but produced an effect which was very large in absolute terms. In

---

[1] This work was supported by Grant G-14223 from the National Science Foundation.

fact, the difference produced by a 50-decibel difference in CS intensity was approximately the same size as that produced by values of .5 and 5.0 psi in the strength of the air-puff UCS. These represent rather extreme values of a variable well known to be highly effective in eyelid conditioning experiments. This large difference was consistent over all values of the other two variables. A rather inescapable conclusion seems to be that something more than statistical precision resulted from the within-*S*s design. Apparently, the exposure of the *S*s to the two intensities of the CS within the experiment served to increase the magnitude of the effect over what is obtained when *S* experiences only one intensity value. This is the case in the usual experiment in which a separate group of *S*s is provided for each value of the variable under study.

What is required to establish this conclusion firmly, is to conduct investigations in which the two methods of studying intensity functions are directly compared. This means, in effect, the comparison of a between-*S*s experiment with a within-*S*s where the same values of the independent variable are used in both. This turns out to be a reasonable, but neglected, experimental design in psychological research. The use of repeated measures on the same *S*s is demanded by some problems, such as the effect of increasing amounts of practice on learning. In other cases, however, within-*S*s designs are used to achieve statistical precision through elimination of individual difference variance, and to effect economies in the number of *S*s required. Such use of the design usually involves, often implicitly, the assumption that the treatment effects obtained are the same as would be obtained with a larger, less economical random groups design. As suggested in the case of stimulus intensity, this is not necessarily so. Even though the values of an independent variable may be the same, the two types of experiment are not investigations of the same problem. Between- and within-*S*s relationships are, in principle, independent, and the laws obtained may or may not be similar. Furthermore, as in the present instance, the difference between the two types of functions may represent a problem of considerable psychological interest. Two experiments have been conducted investigating the effects of stimulus intensity under these different conditions.

EXPERIMENT I

Method

The first experiment was an investigation of CS intensity in eyelid conditioning. Conventional recording and stimulus programing equipment was used (Grice & Davis, 1960). The *S*s were university undergraduate women. The basic parameters of the experiment were: a 1-psi air puff as UCS; a .5-second interstimulus interval; an average intertrial interval of 20 seconds; and a buzzer warning signal at an average of 2 seconds before the CS. The CS was a 1000-cycle tone of .5-second duration, presented in earphones by means of an electronic switch with a 30-millisecond rise and decay time. Its intensity was either 50- or 100-decibel SPL. One group of *S*s (Group L) was conditioned for 100 trials with the loud tone only, and a second group (S) received the soft tone. Two groups (LS-L and LS-S) received 50 conditioning trials with each tone. The order of presentation was random with the restrictions that there be no more than 4 successive trials with one intensity and that there be 10 presentations of each tone in successive blocks of 20 trials. Two orders were used, one the mirror image of the other formed by interchanging the loud and soft tones. Half of the *S*s in each group had one order, and the other half, the other. Procedure for the two LS groups was identical, but for purposes of statistical analysis, the loud tone data were used for LS-L and the soft tone data for LS-S. There were 20 *S*s in each group.

Results

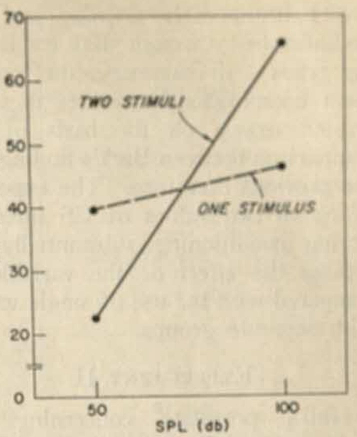


FIG. 1. Percent CRs during last 60 trials to the loud and soft tones under the one- and two-stimulus conditions.

The results are presented and analyzed in terms of percentage of anticipatory CRs. The major findings of the experiment are quickly seen from an inspection of Figure 1. The points connected by the broken line represent percent CRs on Trials 41–100 for the two groups having only one CS. The points connected by the solid line represent the loud (LS-L) and soft (LS-S) tone data for the two groups which received both CS intensities. The magnitude of the intensity effect is more than five times as great for the condition in which *S*s received both stimuli as compared with the single value condition. This interaction is also significant, $F(1, 76) = 8.32$, $p < .01$. The combined intensity effect is also significant, $F(1, 76) = 18.23$, $p < .001$, but for the single stimulus condition alone it is not, $t(76) = 0.98$. Other individual group comparisons are also of interest. For the loud tone, response frequency is significantly greater for the two-stimulus condition than for the one, $t(76) = 2.29$, $p < .05$. The inferiority of the two-stimulus condition on the soft tone does not reach significance, $t(76) = 1.79$, $.10 > p > .05$. In Figure 2 are presented the conditioning curves for the two single stimulus groups and data for the loud and soft stimuli for groups LS-L and LS-S, respectively. Data points are means of 20 trials for Groups L and S and of the appropriate 10 trials for Groups LS-L and LS-S. Basically, these curves show that the phenomena illustrated in Fig-

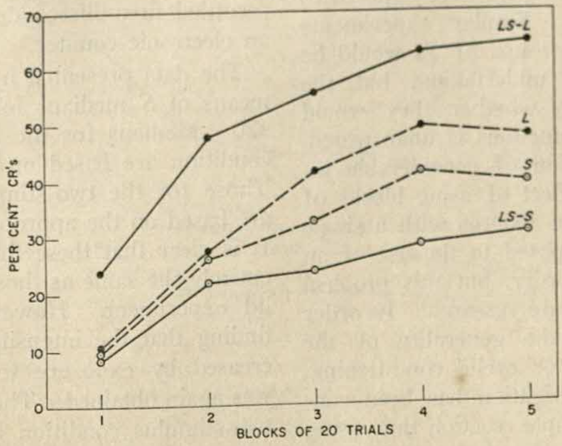


FIG. 2. Conditioning curve data showing percent CRs to each stimulus under each condition of administration during successive blocks of 20 conditioning trials.

ure 1 begin early training and are undiminished through 100 conditioning trials. In summary, the experiment unambiguously verifies the conclusion drawn on the basis of the comparison between Beck's finding and the previous literature. The exposure of *S*s to two values of CS intensity during conditioning substantially increases the effect of the variable as compared with the use of single values with separate groups.

Experiment II

Hull's postulate concerning the dynamogenic effects of stimulus intensity was, in part, based on results from studies of signal intensity in simple reaction time, although he probably did not use the best data available, e.g., Chocholle (1945, 1948). In these experiments clear and orderly functions have been obtained. It is of interest that Cattell's (1886) data cited by Hull, as well as those of Chocholle, were obtained from repeated measures on highly practiced *S*s. Chocholle, for example, would run about 20 trials with one intensity and then shift to another. In other words, both employed within-*S*s designs in which the *S*s were exposed to all intensities. Similar experiments with separate groups of *S*s would be a tremendous undertaking, but the question as to whether they would yield similar functions is unanswered. Another question of considerable interest is the effect of using blocks of trials of various lengths with a single stimulus as opposed to the use of an unpredictable order, but this problem must await future research. In order to investigate the generality of the above finding for eyelid conditioning, a similar investigation has been conducted with simple reaction time using two intensity values of an auditory signal.

The response consisted of pressing a conventional telegraph key. Reaction signals consisted of 1000-cycle tones of 1-second duration, and of 40- and 100-decibel SPL. The 60-decibel separation was used rather than the 50-decibel of the eyelid study because Chocholle's data suggested that a wide separation was necessary in order to obtain substantial effects, if near-threshold values were to be avoided. The foreperiod signal was a buzzer and preceded the tone by 1, 1.5, and 2 seconds in an irregular order. Trials were run at approximately a rate of 4 per minute. There were 120 trials. The experimental groups were analogous to those of Experiment I. Groups L and S always received one signal intensity. Groups LS-L and LS-S received half of the trials at each intensity in an unpredictable order. For Trials 21 to 120, one of the random orders for Experiment I was used throughout. Conventional reaction time instructions were given. The *S*s were women from the same population as for Experiment I. There were 20 *S*s each in Groups L and S, and 10 *S*s each in Groups LS-L and LS-S. Reaction latencies were recorded in milliseconds by means of an electronic counter.

The data presented in Figure 3 are means of *S* medians for Trials 21 to 120. Medians for the single-stimulus condition are based on all 100 trials. Those for the two-stimulus condition are based on the appropriate 50 trials. It is clear that these findings are not entirely the same as those for the eyelid experiment. However, the basic finding that the intensity effect is increased by exposure to both stimuli was again obtained. The slope for the two-stimulus condition is significantly steeper than for the one-stimulus condition, $F(1, 56) = 4.48$, $p < .05$.

Rather than the crossover effect found in the eyelid study, there is an overall slowing of reaction time in the two-stimulus condition. This effect is significant, $F(1, 56) = 16.59$, $p < .001$. This is produced primarily by the response to the soft tone where the difference between the two conditions is significant, $t(56) = 4.38$, $p < .01$. The difference for the loud tone is not significant. The combined intensity effect is significant, $F(1, 56) = 42.97$, $p < .001$, and even the smaller effect for the single stimulus condition alone is also significant, $t(56) = 4.13$, $p < .01$. Practice curves are presented in Figure 4. After an initial improvement, the results are quite stable, and there is no indication that the interaction between intensity and the experimental conditions is decreasing with practice. Thus, in spite of certain differences in detail in these two situations studied, both experiments confirm the general principle that differential responsiveness on the basis of stimulus intensity is increased by multiple stimulus presentation.

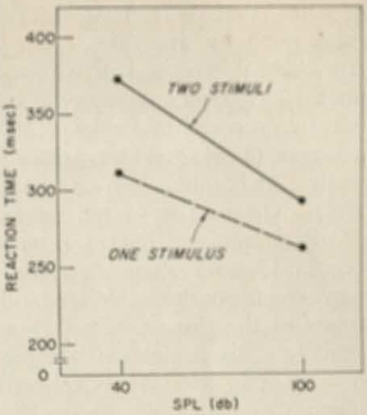


FIG. 3. Means of S median latency on the last 100 reaction time trials.

THEORETICAL IMPLICATIONS

These experimental findings are of considerable theoretical interest. It is at once apparent that Hull's stimulus intensity dynamism postulate will not account for them. He assumed that V was a simple function of stimulus energy, but these experiments clearly show that the dynamogenic property of a given energy depends largely on what other stimuli are presented. It is

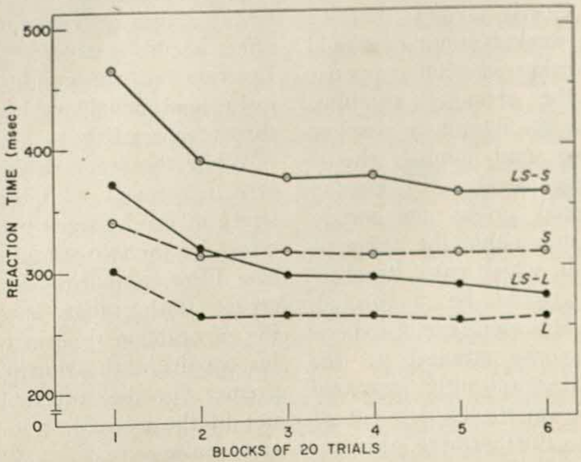


FIG. 4. Practice curves for each stimulus under the two experimental conditions in the reaction time experiment.

clear, then, that there must be no invariant intensity functions.

Another hypothesis, still within the framework of Hull's theory, was independently proposed by Perkins (1953) and Logan (1954), and has been more recently elaborated by Champion (1962). These writers have suggested that the dynamism concept is superfluous in Hull's system, and that intensity effects are deducible from other portions of the theory. This view is essentially a discrimination interpretation. Making use of generalization gradients for excitation originating at the CS intensity value, and gradients of inhibition originating at the background intensity, they were able to deduce greater net excitatory strengths for stronger stimuli. There are some problems involved in the application of this theory to the present situation. If we assume, on the basis of the conditioning curves, that all habit strengths are approximately at asymptotic level, and hence about equal, the theory predicts equal differences between the two stimuli in the two situations. There is one feature of the eyelid conditioning data which is particularly embarrassing to this theory. The theory rather clearly predicts that the addition of a weaker stimulus should result in a weaker *absolute* response strength to the stronger stimulus. This is because the inhibitory gradient from the background intensity should start at a higher level. This prediction would hold unless the strong stimulus was not within the range of this gradient, in which case, its absolute strength would be unaffected. However, in the present eyelid experiment, the response strength to the loud tone was significantly *increased* by the presence in the experiment of the soft tone. Furthermore, this occurred in spite of the fact that it received only half as many reinforced trials as in the single-stimulus situation.

These present data tend to suggest an interpretation in terms of a genuine dynamogenic effect associated with the intensity of the eliciting stimulus. Obviously, however, more than mere stimulus intensity is involved. It appears that some form of contrast effect is an important element in this phenomenon. One approach would be in terms of an adaptation-level concept similar to that of Helson (1959). This interpretation is particularly well able to handle the result of the conditioning experiment. For example, one could make the simple assumption that the dynamogenic potency of a stimulus depends upon its departure from adaptation level (AL) rather than upon absolute intensity. In the case of the single stimulus situation, then, the AL should be near the stimulus value. This would lead to the prediction of minimal intensity effects, since, irrespective of absolute intensity, departure from AL would be small. In the two-stimulus situation, however, the AL might be expected to lie between the two values. With the weak stimulus below AL and the strong above, an exaggerated intensity effect would be expected. The crossed functions of the conditioning experiment lend considerable credibility to this interpretation.

While the reaction time experiment is in agreement with the conditioning study in producing a greater intensity effect for the two-stimulus condition, it also differs in that the functions do not cross. Rather, there is a general slowing of reaction time in the two-stimulus condition, the slowing effect being greater for the soft tone. This does not fit the adaptation-level interpretation quite so neatly. However, it seems possible that, in this situation, the longer reaction times reflect a set phenomenon resulting from the unpredictability of the stimulus. There is little support for this view in the disjunctive reaction time experiments. There, the latency seems to depend upon the amount of information transmitted (Bricker, 1955), and irrelevant information in the stimulus which does not increase information transmitted has no effect, at least at high levels of practice (Morin, Forrin, & Archer, 1961). There seem to be little data, however, on this point for simple reaction time where there is no response selection, and, in this sense, no information transmitted. Mowrer, Rayman, and Bliss (1940) and Mowrer (1941) did, under some conditions, find that the occurrence of an unexpected (high informational) stimulus resulted in a slower reaction even though there was no change in the response required. The present data certainly suggest this, and the matter needs further exploration. This interpretation suggests that the intensity functions would cross except for the presence of an additive constant due to the set variable. The data of Morin, Forrin, and Archer (1961) hint that the crossing might be obtained with extremely extended practice. At least, these considerations indicate that the reaction time data can very possibly be subsumed under the adaptation-level interpretation, if the set explanation turns out to be correct.

Whether or not adaptation-level theory will provide a fully adequate explanation for these phenomena is a matter for future decision. At least a number of experimental approaches are suggested by the findings and by the interpretation. Among the experimental variables which should be of interest are: the number and range of intensity values employed; the relative frequency of occurrence of each intensity used; the length of the trial blocks in which a particular value is consecutively presented; the degree of predictability in the stimulus presentation sequence; the mere exposure of *S* to certain stimuli without reinforcement or requiring a response; and the role of background stimulation. Of course, experimental situations other than eyelid conditioning and reaction time should be investigated. The fact that variable stimulation leads to stronger and more persistent orienting reflexes, also suggests the possible relevance of that phenomenon. It is the intent that the present discussion be limited to situations involving human *S*s. However, there is the possibility that the problem is somewhat different with animals and that stimulus intensity is a more effective variable with them (e.g., see Kimble, 1961, p. 120). If this is true, it may be that attention getting properties of strong stimuli are of more importance with animals than in most human experimentation. Perhaps this will suggest additional approaches to the problem.

A word should be added about the problem of response evocation by stimulus onset versus stimulus cessation. In general, it appears that, for a given intensity, it makes little difference whether the critical event is onset or cessation, so far as response producing potency is concerned. (Champion, 1962; Chocholle, 1948; Logan & Wagner, 1962.) While the discrimination theory of Perkins and Logan will account for this similarity, the doubt cast upon this interpretation by the present data suggests that the dynamogenic property of stimuli depends upon the amount of energy change in relation to *S*s' present condition of adaptation. The contention of Logan and Wagner that a dynamogenic theory implies greater effectiveness of onset is not a necessary part of such a theory. This is an interesting idea, though, since it implies that a sudden decrease in stimulus energy will have dynamogenic properties, if this event is the condition for response evocation. Certainly, the cessation condition requires exploration with the within-*S*s type of experiment as well as the onset condition.

STUDIES OF INTENSITY GENERALIZATION

The present findings have a rather interesting implication for studies of generalization along intensity dimensions. Such experiments inherently involve the presentation of more than one stimulus to individual *S*s. It is thus to be expected that larger intensity effects will be found in generalization studies than in independent group experiments investigating intensity directly. It is apparently no accident that Hull relied so heavily on generalization data in his discussion of the stimulus intensity dynamism, and this chief interest in the concept seemed to be its application to intensity generalization functions. It is well known that generalization tests are conducted in different ways. The *S*s may be tested at only one value, either by a single test trial or by a series of extinction trials. Either of these methods involves a single intensity shift, but the extinction method then involves a series on the new intensity without further change. On the other hand, each *S* may be tested at all intensities, usually in an unpredictable order, counterbalanced between *S*s. There may be more than one cycle of test trials. There is reason for thinking that these methods may not yield the same generalization gradients. The problem raised here suggests that they may not result in the same intensity effects either. Data are not available for adequate predictions, but a likely speculation would be that the counterbalancing procedure would produce the greatest intensity effect, and the extinction procedure the least. The experiment by Hovland (1937), using the counterbalancing method, and that by Hall and Prokasy (1961), using a single test trial, produced reliable intensity effects, while the Grant and Schneider (1948, 1949) experiments, using extinction at a single value, did not.

TABLE 1

ANALYSIS OF HOVLAND'S (1937) DATA ON INTENSITY GENERALIZATION

Source	*df*	*F*
Between subjects	31	
CS intensity (CS)	1	.56
Error	30	(149.90)
Within subjects	352	
Test intensity (TI)	3	27.74**
Cycles (Cy)	2	1.81
TI × Cy	6	.86
TI × CS	3	3.00*
Cy × CS	2	.50
TI × Cy × CS	6	1.60
Error_w	330	(27.25)
Error_w (TI, TI × CS)	90	(26.22)
Error_w (Cy, Cy × CS)	60	(67.91)
Error_w (TI × Cy, TI × Cy × CS)	180	(14.21)

Note.—Error mean squares in parentheses.
* $p < .05$.
** $p < .001$.

The most widely cited experiment using the multiple test procedure is that of Hovland (1937). This experiment is still of considerable interest for several reasons. Briefly, Hovland conditioned 16 *S*s on the GSR at 40-decibel loudness level, and a second group of 16 at 86 decibels. All *S*s were then tested by means of extinction trials at 40, 60, 74, and 86 decibels in counterbalanced orders. Three cycles of test trials were conducted. While appropriate methods of analysis did not exist at that time, all data were published, and a complete analysis is presented here. The basic data, averaged over the three test cycles, are presented in Figure 5, and the summary of the analysis is presented in Table 1. The large effect of test stimulus intensity is apparent in the figure and is highly significant. This large and reliable intensity effect, as opposed to Grant and Schneider's (1949) negative finding for the GSR, is very likely due to the presentation of all stimulus intensities to each *S* as opposed to the Grant and Schneider procedure in which each *S* received only one training and one extinction intensity. This conclusion must be made with some reservation, however, since individual difference variance was not included in this comparison as it was in the Grant and Schneider experiment.[2] The difference in slope between the two functions of Figure 5 may be interpreted as indicating true generalization decrement and is significant (TI × CS), but it accounts for a trivial portion of the variance. Hovland pointed out that the generalization gradient was flat on the first test cycle when averaged over training intensities, and that generalization decrement only appeared on the second and third test cycles. However, it turns out that this change in slope was not statistically significant (TI × CS × Cy).

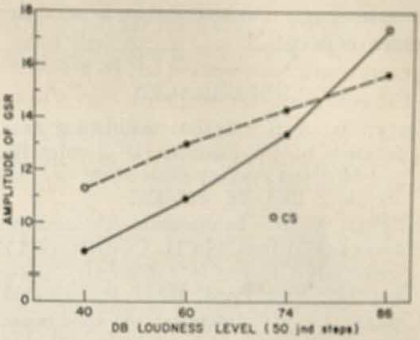


FIG. 5. Test trial data for Hovland's (1937) intensity generalization experiment. (Each point is mean of three test trials.)

Another feature of the Hovland design is that it may be regarded as a test of whether the intensity of the CS during conditioning is a determinant of the associative strength of conditioning. In this respect, it is related to the original purpose of the Grant and Schneider experiments, and to the Spence (1953) experiment of UCS intensity. This experiment, however, involves the counterbalanced presentation of four intensities to each of two CS intensity groups, rather than using a separate group for each test intensity value. The result is consistent with those of Grant and Schneider and also of Hall and Prokasy (1961) in indicating no significant effect of CS intensity. In this connection, it is significant that Hull (1951, 1952) regarded the stimulus intensity dynamism during conditioning (V_1) as a determiner of habit strength. Most of the data do not seem to warrant the conclusion that CS intensity does influence associative strength. However, it appears that while Hull thought of the dynamogenic properties of test stimuli (V_2) as dependent on stimulus energy, he apparently thought of V_1 as dependent only upon the age of the stimulus trace. It now appears that the dynamogenic properties of stimuli, dependent upon intensity and adaptation level, and the degree of associative efficacy, which depends on the interstimulus interval, should not be regarded as manifestations of the same concept.

---

[2] In the Hall and Prokasy (1961) experiment, only data for one test trial were presented. However, eight extinction trials were actually conducted with a single test intensity for each *S*. Unpublished data kindly provided by Hall (J. F. Hall, personal communication, 1963) indicate that the effect due to test stimulus intensity diminished during extinction. From the second trial on, the overall intensity relationship was no longer monotonic, and by the seventh trial, all traces of the effect had disappeared.

REFERENCES

BECK, SALLY B. Eyelid conditioning as a function of CS intensity, UCS intensity, and Manifest Anxiety scale score. *J. exp. Psychol.*, 1963, **66**, 429–438.

BRICKER, P. D. Information measurement and reaction time. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955. Pp. 350–359.

CARTER, L. F. Intensity of conditioned stimulus and rate of conditioning. *J. exp. Psychol.*, 1941, **28**, 481–490.

CATTELL, J. McK. The influence of the intensity of the stimulus on the length of the reaction time. *Brain*, 1886, **8**, 512–515.

CHAMPION, R. A. Stimulus-intensity effects in response evocation. *Psychol. Rev.*, 1962, **69**, 428–449.

CHOCHOLLE, R. Variation des temps de réaction auditif en fonction de l'intensité à diverses fréquences. *Année psychol.*, 1945, **41–42**, 65–124.

CHOCHOLLE, R. Étude de la psychophysiologie de l'audition par la méthode des temps de réaction. *Année psychol.*, 1948, **44–45**, 90–131.

GRANT, D. A., & SCHNEIDER, DOROTHY E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, **38**, 690–696.

GRANT, D. A., & SCHNEIDER, DOROTHY E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39**, 35–40.

GRICE, G. R., & DAVIS, J. D. Effect of concurrent responses on the evocation and generalization of the conditioned eyeblink. *J. exp. Psychol.*, 1960, **59**, 391–395.

HALL, J. F., & PROKASY, W. F. Stimulus generalization to absolutely discriminable tones. *Percept. mot. Skills*, 1961, **12**, 175–178.

HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. *Sensory, perceptual, and physiological formulations.* New York: McGraw-Hill, 1959.

HOVLAND, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291.

HULL, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

KIMBLE, G. A. *Hilgard and Marquis' "Conditioning and learning."* New York: Appleton-Century-Crofts, 1961.

LOGAN, F. A. A note on stimulus intensity dynamism (V). *Psychol. Rev.*, 1954, **61**, 77–80.

LOGAN, F. A., & WAGNER, A. R. Supplementary report: Direction of change in CS in eyelid conditioning. *J. exp. Psychol.*, 1962, **64**, 325–326.

MORIN, R. E., FORRIN, B., & ARCHER, W. Information processing behavior: The role of irrelevant stimulus information. *J. exp. Psychol.*, 1961, **61**, 89–96.

MOWRER, O. H. Preparatory set (expectancy)—Further evidence of its 'central' locus. *J. exp. Psychol.*, 1941, **28**, 116–133.

MOWRER, O. H., RAYMAN, N. N., & BLISS, E. L. Preparatory set (expectancy)—An experimental demonstration of its 'central' locus. *J. exp. Psychol.*, 1940, **26**, 357–372.

PERKINS, C. C., JR. The relation between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, **46**, 225–231.

PIÉRON, H. Nouvelles recherches sur l'analyse du temps de latence sensorielle et sur la loi qui relie le temps à l'intensité d'excitation. *Année psychol.*, 1920, **22**, 58–142.

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

WALKER, EVELYN G. Eyelid conditioning as a function of intensity of conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1960, **59**, 303–311.

(Received April 29, 1963)

Psychological Review
1964, Vol. 71, No. 4, 257-272

THE FUNCTION OF AGGRESSION AND THE REGULATION OF AGGRESSIVE DRIVE

SEYMOUR FESHBACH [1]

University of Colorado

Preliminary to the development of a stimulus-mediating response model of aggressive drive, formal definitions are given of instrumental aggressive acts and aggressive drive-mediated behavior. In defining aggressive drive, a distinction is made between expressive aggression and hostile aggression which is illustrated in the difference between the desire to "hit" and the desire to "hurt." Hostile aggression is assumed to be a learned drive whose primary antecedents are past exposure to punishment and present threats to self-esteem. Several procedures for reducing aggression other than performance of an aggressive response are described and the function of diverse measures responsive to changes in expressive aggression, aggressive drive, and aggressive response strength, particularly when evaluating the consequences of an aggressive act, are analyzed.

The general problems to which this paper is addressed are the functional properties of an aggressive act and the mechanisms by which aggression is reduced. We will be particularly concerned with the following hypothesis which is central to the theoretical analysis presented by Dollard, Doob, Miller, Mowrer, Sears, Ford, Hovland, and Sollenberger (1939) in the frustration-aggression monograph: "The expression of an act of aggression is a catharsis that reduces the instigation to all other acts of aggression [p. 53]." There are three terms or phrases in this hypothesis which require further explication in order to ascertain its theoretical implications and the empirical operations which would be appropriate to a test of this hypothesis. We must establish what constitutes an "act of aggression," specify the properties of an "instigation to . . . aggression," and evaluate the function of "catharsis" in determining the meaning of the above proposition.

AGGRESSIVE ACT: INCIDENTAL, INSTRUMENTAL, AND HOSTILE

A troublesome problem in the investigation of aggressive behavior is the definition of the term aggression. The most formal attempt at a rigorous definition of this concept has been made by Buss (1961), but, as has been indicated elsewhere (Feshbach, 1961), important difficulties still remain. There is general agreement (Berkowitz, 1962; Buss, 1961; Dollard et al., 1939) that aggressive behavior denotes injury to some person or object. Disagreement arises in regard to the inclusion of "intent" or some other motivational construct in the criterion for an overt aggressive act. Insistence upon exclusion of motivational considerations in categorizing an injurious act as aggressive results in several difficulties. In particular, accidental injury and the unintended infliction of pain (e.g., the immediate pain consequent to the dentist's injection of novocain) would be classified as aggressive while an intended injurious response that fails to achieve its goal would then fall into the nonaggressive class. In addition, there are innumerable instances in a competitive society in which injury to some party is an inevitable consequence of the norms governing the distribution of economic and prestige rewards in the society. To vote for Candidate A is to injure Candidate B although, to be sure, voting for Candidate B might result in a much greater injury to the population at large. Firing an employee who needs and wants the position inflicts injury upon him although the firing be justified by the economic advantages to be gained by hiring cheaper or more competent labor. The assertion of one's own interests at the cost of a competitor's or colleague's interests results in a loss to the competitor regardless of the legitimacy of one's own claim. Social conventions attest to the difference between competition-produced and aggression-instigated injury and loss. However, it must be noted this discrimination is not easily made and it has been observed that many members of our culture apparently generalize the inhibitions evoked by aggression-produced cues to cues stemming from the assertion of their own self-interest. This equivalence is evidenced in the several usages of the adjective "aggressive" which carries connotations both of assertiveness and hostility.

---

[1] The author wishes to express his gratitude to Norma Feshbach, Francis Irwin, and Paul Mussen for their critical reading of the manuscript and their helpful comments.

An initial draft of this paper was prepared at Stanford University while the author was the recipient of a United States Public Health Research Fellowship.

For these reasons, it is useful to restrict the term aggression to those aggressive acts which are intended, or at least distinguish between unintentional and motivated aggression. Unintentional aggression is defined as an act which, although resulting in injury, is not contingent upon its injurious consequences. Intentional aggressive acts display the formal properties of a preference as defined by Irwin in his analysis of the relationship between discrimination and preference (Irwin, 1958). Not all intentional aggressive acts are functionally equivalent, however. Thus the distinction is made between *instrumental* aggression which is directed toward the achievement of nonaggressive goals, and *hostile* aggression, or aggressive drive for which the goal response is injury to some object (Sears, Maccoby, & Levin, 1957). Irwin's definition of "desire" and his demonstration that preference is not equivalent to desire is pertinent here (Irwin, 1961). Hostile aggression is motivated by a desire for a noxious outcome while the term instrumental aggression shall be restricted to describe behaviors producing injury which are motivated by a desire for some outcome other than injury to an object. Such outcomes may include reinforcement of motives varying from attention-getting to self-preservation.

There is an increasing tendency in contemporary theoretical treatments of aggression (Bandura, Ross, & Ross, 1961, 1962; Buss, 1961) to stress instrumental aggression and either discount the hostile, aggressive drive concept or relegate it to secondary importance. The noxious consequences of an aggressive act are viewed as one of a large number of possible reinforcers, and the analysis of aggression presumably need concern itself only with the reinforcing stimulus conditions and the discriminative stimuli eliciting the response. This analysis seems to provide a satisfactory account of many forms of aggressive interaction in infrahuman species (Scott, 1958) and has been applied with some success to aggressive behaviors in young children and adolescents (Bandura & Walters, 1959; Buss, 1961). However, it fails to do justice to the cultural implications of an aggressive act and particularly to its interpersonal properties. It will be contended here that the injurious outcomes resulting from instrumental aggression must be taken into account when considering the variables which influence the aggressor's behavior.

The prototype of aggressive behavior in animals and young children is fighting and biting. An 18-month-old child, when thwarted, may attack the frustrating source by biting the object or by kicking or hitting it. If the frustration is eliminated—the child, for example, regains a toy that has been taken away—then these attacking responses are presumably reinforced. On the next occasion, then, in which the child experiences a similar frustration, he is likely to respond in a similar manner. But society cannot tolerate for very long this simple manner of problem solving and the culture requires the child to make a complex set of stimulus and response discriminations. On the response side, he quickly learns that biting is taboo, that kicking is unmanly, and that fighting must be carried out solely by the hands, unaided by weapons. If he is a powerful youngster, he may have to learn to modulate the vigor of his attack. On the stimulus side, he may be punished for striking his parents, but rewarded for fighting his friends (Bandura & Walters, 1959). He has to discriminate not only between objects toward whom he can express aggression without disapproval or punishment, but also the occasions for which an aggressive response is appropriate. He may be taught to defend himself but not to initiate an attack. Most difficult of all, he must learn to distinguish between intentional and unintentional frustration and injury. The regulation of this complicated repertoire of behaviors is effected through a variety of rewards and punishments which the parents may sometimes inadvertently administer and indirectly through identification with and imitation of the parents and other models.

Many of the details of this process have been traced elsewhere (Bandura & Walters, 1963; Buss, 1961; Sears, Maccoby, & Levin, 1957). Of particular relevance to the present discussion are features of the regulation of aggression which pertain to the immediate noxious consequences of an instrumental aggressive act. Whether a child achieves or fails to realize some goal through an aggressive act, his aggression will be labeled. He will be told that he is hurting someone, that he is inflicting pain, and he learns to associate this label with the reaction of the object and with his own aggressive response. The individual who crudely sweeps others aside in his single-minded pursuit of a goal and who remains ignorant of the destruction and havoc he wreaks, is a rare person indeed.

In brief, responses producing the cues of "it hurts" in others will lead to reward, nonreward, to punishment or to some combination, as the case may be. The cues can then serve to mediate that very diverse class of behaviors, that we label aggressive.[2] It is recognized that physical aggression, verbal attacks, humiliation of others, rejection, etc., may be differentially reinforced. However, to the extent that there is functional equivalence between these radically different behavioral forms in that preference for their use or avoidance in achieving nonaggressive goals is correlated, their common consequence is probably an important element in determining this set of preferences.

---

[2] The cues of pain in others, as Sears (1958) has suggested, may become secondary reinforcers and serve to maintain aggressive responses in the absence of other rewards. This possibility will be considered in the discussion of aggressive drive.

The noxious consequences of an aggressive act may exert still another kind of influence on the use of instrumental aggression. Early in their developmental history, children demonstrate affective reactions similar to those evoked in someone whom they are witnessing (Murphy, 1937). This response pattern is especially evident when children are seen watching dramatic films. It has also been noted in daily interactions with peers. A child may become anxious and distressed simply by observing an act of aggression directed toward another child. The first child responds vicariously to the experiences of the second.[3] Observation by an empathic child of the consequences of an aggressive act directed toward another child should tend to elicit distress responses in the child observer even if he himself is the instigator of the aggressive act. One would expect these painful, vicarious responses to function as inhibitors of the child's own aggressive responses. These inhibitory tendencies should occur even if the instrumental aggressive act is successful in achieving some nonaggressive goal. Thus children high in empathy should, on the whole, manifest less overt aggression than children low in empathy.[4]

Although there is a dearth of relevant empirical data, informal, clinical observations suggest that empathic inhibition often plays a significant role in restraining the infliction of pain on others. Interesting psychological questions arise in connection with the genetic antecedents of this process and with the dynamic, situational factors which attenuate and exaggerate its potency. In regard to the latter, similarity between the observer and the victim should facilitate its potency. In aggressing against another person, it is convenient to perceive him as someone very different and, if at all possible, to dehumanize him. Shylock's famous soliloquy, affirming his membership in the human species, is an attempt to undo this differentiation and thereby reduce aggression directed toward him. The immediacy and intensity of the pain should also be relevant factors. The discomfort produced by physical attack is direct and apparent, the pain produced by rejection is distal and more complexly determined. The parameters of empathic inhibition may help explain why physical punishment for criminals is frowned upon in modern penal systems, although other extremely punitive devices still persist. In the hedonic calculus employed by the prisoner, 20 lashes may be preferable to 10 days of isolation. A further factor in discouraging the persistent use of physical punishment is its possible "brutalizing" influence upon the punishing agent. While the aggressor is motivated by empathic responses to inhibit aggressive acts, if, for other reasons, he continues to perform these aggressive acts, the empathic-produced pain may result in the evocation of responses which are incompatible with the empathy response. When a father apologetically tells his son while spanking him, "this hurts me more than it does you," his self-report may be veridical for the initial spankings, but if this mode of punishment is frequently used, the empathic response is likely to be weakened. Again, informal observation suggests that continual exposure to human pain and suffering will result in the acquisition of defensive behaviors—e.g., intellectualization—which are incompatible with the empathy experience.

---

[3] This correlation alone is not sufficient to demonstrate a genuine empathy response. Sources of pseudo-vicarious instigation have been reviewed in detail by Berger (1962).

[4] This straight-forward deduction from the assumption of empathic inhibition has yet to be directly confirmed. In Lois Murphy's (1937) study of sympathy in nursery school children, a positive correlation was obtained between teachers' ratings of aggressive behaviors and ratings of sympathetic behaviors which are similar in a number of instances to indexes of empathy. This positive relationship cannot be taken as strong evidence against the hypothesis since, as the author suggests, both aggressive and sympathetic responses are instances of socially oriented behavior. The positive correlation in this age group, then, appears to be largely a function of the social maturity of the child and the extent to which he participates in social interaction.

To summarize, it is proposed that even in so-called "instrumental aggression," the noxious consequences of an aggressive act influence the behavior of the aggressor. Moreover, as the psychoanalysts are fond of arguing, much of human behavior seems overdetermined. The boy who strikes his younger sibling to get his mother's attention may well obtain a small quantum of satisfaction from his sibling's discomfort as well as the gratification attendant upon his mother's presence and interest. The remaining discussion will be concerned with aggressive behaviors for which the noxious consequences are the primary goal.

INSTIGATION TO AGGRESSION

The frustration-aggression hypothesis asserts that frustration of an ongoing activity produces an instigation whose goal response is injury to some person or object. This conceptualization retains most of the dynamic properties of the psychoanalytic theory of aggression without the limiting and unnecessary assumption of an instinctive drive. The strength of the instigation is directly linked to the frequency and intensity of frustrating experiences and in contrast with an instinct assumption, data can be obtained bearing upon the utility of this hypothesis. Criticisms of the frustration-aggression hypothesis have been largely concerned with the invariance of this relationship and with the occurrence of nonaggressive responses to frustration. Insufficient attention, however, has been given to the "instigation" construct and particularly to the nature of the goal response. Although Dollard et al. (1939) in their original formulation used the concept of instigation in a rather broad, general sense to refer to the antecedent conditions correlated with an aggressive response, the use of the term "goal response" suggests a correspondence between the notion of drive and instigation. Also, the extent to which learning enters into the frustration-aggression relationship is left ambiguous. It was sufficient for the author's purpose at that time to state this hypothesis regardless of its origins. However, in order to derive means of controlling and reducing aggressive drive, further analysis is required

In considering the role of learning it is helpful to distinguish between the rage or anger response and the motivated infliction of injury. Sears, Maccoby, and Levin (1957) have noted that anger "appears to be an instinctive response to certain kinds of restraint and discomfort" while the desire to hurt someone in the process, however, is not so clearly instinctive. This distinction is essential to the analysis of the dynamics of anger and aggression that is presented in this paper. All emotions have expressive components and most emotions appear to have a directive component. In the case of anger in infants, the expressive component includes changes in facial appearance, the emission of loud sounds and agitated, vigorous movement of the limbs (Goodenough, 1931). This pattern of behavior is assumed to be an innate response in the human infant to frustration. As the infant acquires greater motor control and perceptual sensitivity, these external movements become less diffuse and more specifically directed toward the perceived source of frustration. These aggressive acts will on occasion have instrumental value in removing the frustrating agent, and, to the extent that the behavior is rewarded, the innate connection between frustration and the anger reaction will be strengthened. To paraphrase the frustration-aggression statement, in this early state of development (from infancy through approximately the second year), frustration produces an instigation to hit rather than to hurt; the topography of the response is more important than its noxious consequences. Given a frustrating agent and a choice of three alternatives such that the child can respond with a vigorous, aggressive motor response with no pain feedback from the object, or can inflict injury upon the object by a mild motor response such as pulling a lever, or witness someone else inflict injury upon the object, the 2-year-old child should prefer the first alternative.[5]

The relative neglect of the expressive component of anger in comparison with its destructive aspects is understandable in view of the prevalence of hostility in human relations. However, its functional significance was discussed in detail by Darwin almost a century ago. Darwin (1873) noted the threat function of an aggressive display and his first principle of "serviceable associated habits" is concerned with emotional expression or remnants of action that once were of direct use to the organism. Of more interest to us is his third principle of actions due to the constitution of the nervous system. Here he views emotional expression as an overflow of energy. The excited motor activity of the enraged animal serves as a means of tension regulation and discharge. There are striking points of similarity between Darwin's third principle and Freud's Pleasure Principle. It will be recalled that Freud postulated that instinctual impulses seek immediate motor discharge. It seems reasonable to assume that the expressive responses associated with anger arousal serve this function. It should be noted that a significant change has been introduced in the character of the instinctual impulse of aggression as Freud conceived it. To the extent that it is meaningful to speak of discharge and release of an *instinctual* aggressive impulse, the index of release is affective expression and not injury to some person or object.

As part of the socialization process, one learns to inhibit affective expression as one is taught to inhibit destructive behavior.[6] The child learns to substitute language and fantasy for some forms of overt motor expression (Rapaport, 1952) and the development of thought processes enables him at the same time to delay impulse expression. Acknowledgment, rehearsal, and the verbal communication of feelings (aided, as a rule, by facial gestures) take the place of the motoric expression of emotional tension. The sensoritonic theory as formulated by Werner and Wapner (1952) assumes a vicarious relationship between motor inhibition and cognitive expression of tension. The results of several studies concerned with the relationship between motor inhibition and Rorschach movement responses are consistent with this assumption (Meltzoff, Singer, & Korchin, 1953; Singer, Meltzoff, & Goldman, 1952). The writer has reported an experiment in which subjects, encouraged to deny their feelings and then exposed to a fear-arousing situation, projected reliably more anxiety onto a stimulus person in a film than a comparable group of subjects who had been encouraged to express their feelings (Feshbach & Singer, 1957). In a more recent study concerned with the restraint of positive feelings (Feshbach, 1963), it was observed that subjects instructed to refrain from communicating a successful achievement through verbal or facial means rated themselves as reliably more happy and attributed more happiness to others than a similarly successful group who were not exposed to this restraint condition. The latter, in fact, hardly differed from a nonrewarded control group. While these studies are by no means conclusive, they are consistent with the assumption of a vicarious relation between affective inhibition and cognitive activity such that cognitive expression of affect may serve as a substitute for motor expression of affect.

---

[5] The relationship between instrumental, expressive, and hostile aggression may be summarized as follows: Using Irwin's (1961) terminology, all three types reflect aggressive preferences. Instrumental aggression is motivated by a desire for nonaggressive outcomes; hostile aggression is motivated by a desire for aggressive (noxious) outcomes while expressive aggression is motivated by a desire for a particular form of responding. Although Irwin's analysis of desire was restricted to response outcomes, it can also be usefully extended to response modes providing that the choice of response mode has some instrumental function.

[6] In their discussion of the development of aggression, Sears, Maccoby, and Levin (1957) observe that children seem to experience relief by acting out aggressive impulses even though other needs may not be gratified. They suggest that the satisfaction and relief comes from the reduction of conflict

The problem of substitute value will be examined in the context of the discussion of catharsis and the reduction of aggressive drive. Before so doing, we have yet to consider the most striking aspect of aggressive drive, namely an organized pattern of behavior for which the goal response is the infliction of injury. Having rejected the notion that this destructive impulse is innate either in the Freudian sense of an instinctual drive or as an unlearned reaction to frustration, the problem arises of accounting for its origins. We shall be particularly interested in the evocation of hostile impulses as a reaction to frustration and threat, recognizing that there are other sources of aggressive drives, e.g., erotic impulse gratification which psychoanalytic theory suggests is a major element in sadism.

Sears (1958) has proposed that the motive to injure others becomes acquired through a process of secondary reinforcement. Successful instrumental aggressive acts result in the elimination of frustration and also in the evocation of pain responses from the frustrator. Because of the consistent association of the perception of pain in others with removal of frustration, the former acquires secondary reward value. While this analysis has been extended by Sears and his collaborators to account for aggressive behavior mediated by an acquired aggressive drive which is independent of frustration, we shall limit our discussion of this developmental account to the simpler situation in which an injured organism finds satisfaction in inflicting pain upon the frustrating source.

There is at the present time no good

---

stemming from the inhibition of aggression. Their hypothesis is similar to the affect-discharge formulation and from an operational standpoint would be very difficult to distinguish from it.

evidence to support or refute the secondary reinforcement explanation of the origin of the impulse to injure others. One can, however, point to several difficulties presented by this particular theoretical account which assumes that the infliction of pain results in the attainment of the frustrated goal, not invariably, to be sure, but at least with sufficient frequency for the reinforcement to generalize to the cues immediately preceding it. One may question the extent to which this assumption holds for most young children. Rather, hurting one's parents or one's older siblings is likely to be met with punishment and also with continued frustration of the blocked goal. An important exception to the latter would be negative attention-getting behavior which may be reinforced even though the attention may take the form of punishment of the child. Attention, which is itself a derived motivation, would then be made to carry the reinforcement burden. A more removed but relevant difficulty is the observation made by students of animal aggression that aggressive behavior in species other than homo sapiens is almost always instrumental to the achievement of nonaggressive goals (Scott, 1958). Animals do not ordinarily attack or kill other animals in order to elicit the cues of pain despite the frequency of successful instrumental aggression associated with some species and with some dominant organisms within a species. The derivation of satisfaction from eliciting pain and suffering in others is a peculiarly human phenomenon, suggesting that we look to factors that are uniquely human as the source of this mode of satisfaction.

One such factor is the child's ability to label and comprehend the contingent events in his environment. We have previously noted that the noxious consequences of a child's aggressive acts are usually labeled as such. When he is punished for the expression of aggression, as well as other misdemeanors, he learns to associate the infliction and perception of pain in others (produced by his own aggression) with the experience of pain in himself. Moreover, these are crudely correlated such that the greater the amount of pain inflicted by the child, the greater the degree of punishment received. Young children, as it were, are subjected to the "lex talionis." More than fear is acquired when the child is punished. He also observes and imitates the behavior of the punishing agent (Bandura, Ross, & Ross, 1961, 1963). Through a process of modeling, he learns a norm or law which essentially states that the infliction of pain is the response made to the experience of pain. He learns to apply this law to certain classes of "infliction of pain" which are categorized as "intentional" and "unfair." He learns that it is appropriate, even required, to retaliate —that is, to inflict injury, under certain social circumstances. The cues of pain elicited in the source of frustration signify that the requirements of the situation have been met. It is not enough to hit; one must also hurt.

This is not to imply that the learned retaliatory behavior pattern is completely unrelated to the "natural" response to frustration and provocation The anger evoked in such situations serves to exaggerate the response strength of these primitive retaliatory tendencies. In addition, the expressive responses associated with anger can be readily utilized to elicit pain cues in others. Finally, the affective anger response may become part of the stimulus complex which initiates the motivation to inflict injury. However, despite the likelihood that anger and aggressive drive (hostility) will be linked together, this need not be the case. The motivation to injure others—or aggressive drive—which can be influenced by anger is, according to the present theoretical analysis, an acquired motivation based upon the internalization of a standard which the child acquires from exposure to particular cultural values and from his own concrete experiences. Thus this analysis directly links the motivation to injure others to the degree of punishment which the child has received, especially for the expression of aggression. There are a substantial numbers of findings consistent with this hypothesis (Bandura & Walters, 1959; Sears, Macoby, & Levin, 1957; Sears, Whiting, Nowlis, & Sears, 1953). There are, however, several alternative explanations of the correlation obtained between punishment and aggressive behaviors, and more precise measures, particularly of the reinforcing properties of the infliction of pain, are necessary in order to prove evidence more specific to this punishment model hypothesis. An adequate test of this hypothesis would require separate indexes of these reinforcing effects and the strength of instrumental aggressive responses. We would expect that individuals who have had a history of experiencing retaliation and counteraggression when they aggress, in comparison to individuals who have a history of reward for aggressive responses, would be more likely to be reinforced by the suffering of a provocateur and would be more likely to persist in a sequence of responses which can be anticipated to culminate in the infliction of injury to the provocateur.

The hypothesis that children acquire aggressive drive through example and percept may account for retaliatory tendencies, but only partially explains the "satisfaction" derived from the perception of pain in one's tormentors. The rewarding properties of this perception may also arise from its status and power implications. We have attempted to draw a sharp distinction between aggressive acts which are instrumental to the attainment of nonaggressive goals and aggressive acts which are motivated by the intent to inflict injury upon some person or object. This distinction becomes blurred when hostility is elicited by a threat to self-esteem. And in human beings, much, if not most occasions for hostility fall into this category, some authors, in fact, considering threats to self-esteem as the primary source of aggression (Worchel, 1960). Thus, the failure to receive an expected invitation to some gathering implies more than interference with a potential pleasurable experience. A student may be annoyed and disappointed at receiving a failing grade. But if he should discover that the instructor is pleased by his failure, he is likely to become furious. A week's solitary confinement in a narrow cell with room barely wide enough to sit is a tremendously frustrating and painful experience. Prisoners become angry and bitter when subjected to this experience as a punishment. But the same experience when endured in connection with voluntary participation in a research project designed to assess the psychophysiological effects of confinement in a space capsule will elicit far less hostility. The difference in these situations is in the implied devaluation and impotence of the frustrated individual. Successful retaliation changes his status relation vis-a-vis the perceived source of devaluation. He has demonstrated that he is not helpless and impotent. Moreover, he need not be responsible for the hostile act. The initial instigator by virtue of being subjected to painful experiences matching those he inflicted on the victim is reduced to the victim's status. The nature of the devaluation and the degree of aggression anxiety would determine the relative preference for self-initiation of an aggressive act versus infliction of injury upon a disliked person by some other individual. That the latter can be effective is indicated by the Rosenbaum and DeCharms study (1960) in which the level of aggression of Subject A, who had been attacked by B, was reduced by overhearing Subject C attack B. The mediating process, however, is unclear in that the reduction in aggression in this study may have been due to the support of A implied by C's attack with consequent reduction of the threat to self-esteem rather than due to the pain experienced by B.

The postulated structural relationship between threats to self-esteem and satisfaction from the infliction of injury in conjunction with the acquisition of the pain-infliction punishment-retaliation norm are proposed as the primary source of the reinforcing properties of the perception of pain in others. We have argued that these variables play a more significant role in determining preferences for behavioral alternatives which elicit pain responses in others than the association of such stimuli with primary reinforcement. It has also been maintained that the relation between anger and the motivation to inflict injury upon others is neither intrinsic nor arbitrary and that the expression of anger is not equivalent to but falicitates aggressive drive-oriented behavior. In addition to the question of the antecedents of aggressive drive, this analysis also bears directly upon the problem of the reduction of the "instigation to aggression."

CATHARSIS AND THE REDUCTION OF THE INSTIGATION OF AGGRESSION

In order to delineate distinctions between what is reduced or modified, we shall adopt the following terminological usage which is similar in a number of respects to that suggested by Buss (1961): (*a*) The terms, hostility, hostile impulse, and aggressive drive shall be used as equivalents; hostile behavior denotes responses which are primarily directed toward the infliction of injury upon some other person or object. (*b*) Anger refers to mediating affective responses whose expressive aspects have already been described. The intensity of anger is usually correlated with the vigor of the hostile response and with the degree of injury which the aggressor is motivated to inflict upon the object. However, expressions of anger, though they may result in injury to some person or object, are not necessarily hostile in intent. Also, hostility may be strong although the affective response is short-circuited or blocked. (*c*) Aggression refers to the commission of acts which produce injury in others where the injury is instrumental to the achievement of nonaggressive goals or is itself the principal goal.

The anger-hostile impulse contellation is elicited by a variety of events broadly labeled frustration which include interference with ongoing instrumental activity, verbal and physical attack, and threats of self-esteem. The reduction of anger and aggressive drive can be accomplished through (*a*) injury to the frustrating source or some form of displaced aggression; (*b*) facilitation of mediating responses which are incompatible with anger and hostility; (*c*) modification of the initial eliciting stimulus

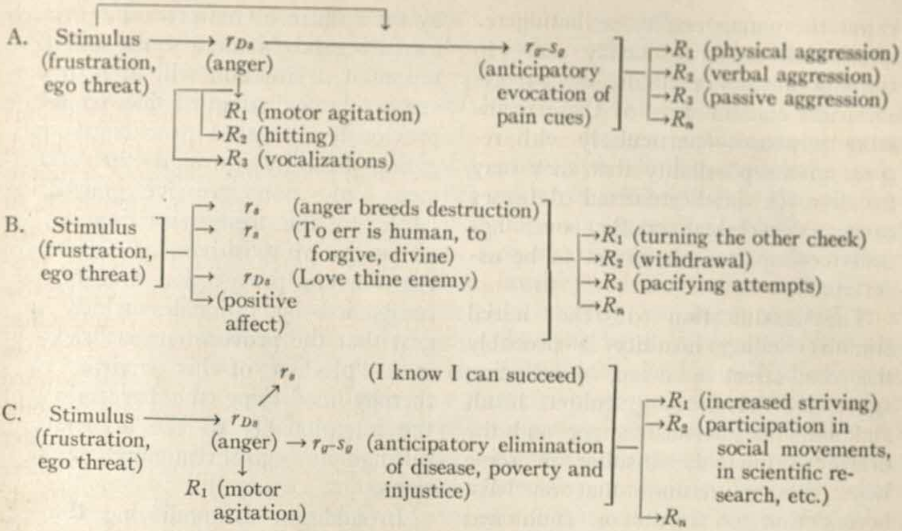

A.
Stimulus (frustration, ego threat)
r_{D_a} (anger)
R_1 (motor agitation)
R_2 (hitting)
R_3 (vocalizations)
r_g-s_g (anticipatory evocation of pain cues)
R_1 (physical aggression)
R_2 (verbal aggression)
R_3 (passive aggression)
R_n
B.
Stimulus (frustration, ego threat)
r_s (anger breeds destruction)
r_s (To err is human, to forgive, divine)
r_{D_s} (Love thine enemy) (positive affect)
R_1 (turning the other cheek)
R_2 (withdrawal)
R_3 (pacifying attempts)
R_n
C.
Stimulus (frustration, ego threat)
r_s (I know I can succeed)
r_{D_a} (anger)
r_g-s_g (anticipatory elimination of disease, poverty and injustice)
R_1 (motor agitation)
R_1 (increased striving)
R_2 (participation in social movements, in scientific research, etc.)
R_n

FIG. 1. Schematic representation of alternative mediating responses to "aggressive" provocation.

condition either through removal of the stimulus or through a change in the meaning of the stimulus. Categories *b* and *c* represent nonaggressive methods of "reducing" aggressive drive.

An example of Category *b* is the advocacy and acquisition of positive affective responses to the stimulus conditions which ordinarily elicit hostility as in the exercise of Christian love and charity toward one's enemy. It may also be possible and, in the opinion of some, desirable to accomplish a similar end while maintaining the anger response, by strengthening nonhostile responses to anger (Kaufmann & Feshbach, 1963a, 1963b). Stratton (1923), some 40 years ago, stated and eloquently argued for this alternative. He notes the undirected character of anger—

> anger is thus seen to be a highly unspecialized compulsion; it has no immediate interest of its own, but arises to protect and to further any interest that you feel [p. 254].

He then suggests that anger can be utilized in the service of socially constructive pursuits.[7] This is what William James (1917) presumably had in mind in his advocacy of a "moral equivalent of war."

These alternative possibilities are diagrammed in Figure 1. The complete short-circuiting of anger and the attachment of nonaggressive mediating responses directly to the instigating stimulus is depicted in Figure 1B while examples of nonaggressive responses to anger are suggested in Figure 1C. The statements in the parentheses adjacent to the mediating stimuli are intended only as shorthand descriptions of the functions of the mediating process and are not meant to imply that these processes are conscious or solely cognitive.

---

[7] Stratton (1923) graphically illustrates the dynamogenic properties of anger in the following quote from a nineteenth century California preacher: "Although I weigh one hundred and twenty pounds, when I am mad, I weigh a ton."

While the nonaggressive mediating responses should theoretically serve to mitigate hostile reactions, the precise empirical consequences of these mediating responses—particularly with respect to the possibility that they may function as intellectualized defenses or may exacerbate conflict over aggressive impulses—have yet to be ascertained.

The modification of the initial stimuli eliciting hostility is probably the most effective means of reducing aggressive drive. A gratuitous insult and snub may arouse anger and the desire to hurt the insulter in some way. Let us assume that one discovers that no insult or snub was intended, that one had misinterpreted the presumed "insulting" comments. Under these circumstances the anger and hostility should disappear. Those operating within the framework of a psychoanalytic drive model might argue that the hostility has been inhibited or repressed and that it must be expressed in order for the drive tension to be reduced. The stimulus-mediating response model which has been presented here takes a simpler view of the matter. The stimulus situation has fundamentally changed and this new meaning evokes a new response. There is no need to satisfy aggressive drive because the stimulus conditions which provoked hostility are no longer present. Changes in the stimulus are frequently accomplished through changes in mediating responses, such that new perceptions of the same situation result from the attachment of different mediating labels, e.g., the classification of a previously provocative situation as "unintentional" or the act of an "irresponsible" agent. As part of the socialization process, such labels become discriminative cues to which nonhostile responses have been fostered by the culture. This process by which mediating labels modify the interpretation of a situation will in many instances be indistinguishable from the previously discussed process of mitigating hostility through the attachment of nonaggressive mediating labels to the same situation. The advocacy of positive, nonaggressive reactions to provocation is often accompanied by rationales which suggest that the provocateur is "sick" or is a "product of his environment," thereby modifying the perception of the intentionality of the act and reducing its ego-threatening implications.

In addition to modifying the external stimulus conditions, the impact of the noxious stimulus can also be altered by modifying the vulnerability of the insulted individual. If the threat to self-esteem is subsequently reduced through a procedure in which the initial instigator is not involved, the insulted person should still become less hostile to the instigator. Evidence consistent with this expectation has been reported by Worchel (1961) and Magaziner and Feshbach (1961). Changes in self-esteem often accompany the expression of hostility and complicate the evaluation of the aggressive drive-reducing effects of the infliction of injury (Rosenbaum & DeCharms, 1960; Thibault & Coules, 1952). We distinguish here between any changes in self-esteem which may be an intrinsic consequence of the infliction of injury and modifications in esteem which are independent of any pain experienced by the object of aggression.

These complicating effects of self-esteem are most apparent in discussion and studies of catharsis, which are further complicated by the variety of meanings which have been attributed to this concept. When the term was initially used by Freud and Breuer (1948) in their papers on the treatment of hysteria, it referred to the tension-reducing consequences of emotional expression: in the case of the affect-drive system of anger-hostility, giving vent to one's hostile feelings was thought to reduce the intensity of the hostility even if no damage had been inflicted upon the object of the hostility. The term catharsis is sometimes restricted to verbal expressions of affect while abreaction is used to refer to motoric affective expressions. In the quotation from the frustration-aggression volume at the beginning of this paper, catharsis denoted a reduction in aggressive drive consequent to the infliction of injury.

These two usages of catharsis[8] correspond to the distinction which has been made between expressive aggression and hostile aggressive drive. The acts of aggression that have been varied in experimental studies of catharsis usually have both an expressive aspect and a noxious consequence although some primarily fall into the former category (Feshbach, 1955; Siegal, 1956; Singer & Rowe, 1962), while others fall predominately in the latter (Berkowitz, Green, & Macauley, 1962; Hokanson & Burgess, 1962; Hokanson & Shetler, 1961).

Apart from the fact that studies of catharsis have not been designed to separate the effects of the expressive and noxious components of an aggressive act, data bearing upon the drive-reducing effects of some direct or indirect aggressive behavior may well yield conflicting results because of the complex consequences of aggressive behaviors (Berkowitz, 1958, 1960). Thus, one implication of the present stimulus-mediating response analysis of anger and hostility is that participation in aggression—whether directly or vicariously—will have no anger or hostility-reducing effect unless anger or hostility has been aroused at the time the aggressive act is performed (Feshbach, 1961). Witnessing aggressive sequences in films or participating in a boxing match cannot be expected to provide a vicarious expressive or hostile function without the organism being aggressively motivated. If anger-hostility is not aroused, then an increase in subsequent aggression may take place as a result of reinforcement of the aggressive response, reduction in aggression-anxiety or aggressive response-eliciting properties of aggressive stimuli (Bandura, Ross, & Ross, 1961, 1962; Feshbach, 1956; Kenny, 1953; Lovaas, 1961).

In addition, although aggressive drive may be reduced by an aggressive act, this effect may be obscured by reinforcement of the aggressive response, by lowered inhibition, and by more subtle influences such as dislike induced possibly by cognitive dissonance (Berkowitz, Green, & Macauley, 1962), so that some measures of aggression may reflect an increment in aggressive behavior. If the measures of aggression one utilizes are about equally responsive to changes in drive strength, then the more similar the response assessed in the dependent measure is to the rewarding, drive-reducing aggressive act, the more likely one is to find evidence of an increment of aggression. Dissimilar measures, in comparison, will be more likely to reflect decrements in aggression. In order to evaluate hypotheses regarding the consequences of an aggressive act, it becomes essential to utilize a variety of measures which are differentially responsive to the diverse effects which have been reviewed above. For example, in a recent unpublished study carried out by the author, angered subjects who were given the opportunity to express their feelings in a sympathetic interview situation, in comparison to a control group, showed a significant increment in aggressive word associations, but at the same time, a significant decrement in displacement of aggression toward an innocent target. Thus in evaluating the effects of an aggressive activity upon subsequent aggression, opposite inferences can be drawn depending upon the measures of aggression used. The problem is not merely one of quantitative differences in the reliability or sensitivity of the measures employed, but rather of qualitative differences in the variables which are being assessed. One purpose of this paper has been to clarify the differential properties of these variables by means of the formal distinctions which have been made regarding the nature of aggressive drive and the relationship between aggressive drive and instrumental aggression.

---

[8] We omit here any reference to the original Greek meaning of catharsis which connotes

> the state of feeling produced by the dramatic tragedy. It meant the stillness at the center of one's being which came after pity or fear have been burned out. The soul is purified and calmed, freed from the violent passions [Scheer, 1961, p. 520].

It is recognized that these theoretical distinctions place rather severe requirements upon the measuring instruments that should be used in studies of aggression. The difficulty in separating the drive and habit components of a response is nowhere more acute than in the case of aggressive behavior. Expressive, hostile, and instrumental functions are interwoven in most aggressive acts and these functions are likely to be further obscured by inhibitory factors. Nevertheless, it is possible through appropriate operations to disentangle these properties. Such an approach is necessary, we believe, to provide a meaningful response to the issue which is central to this discussion—namely, the functional value of an aggressive act—and to intimately related applied social issues which include questions concerning the effects of violence in television and other forms of mass media, the role of social prejudice and intra- and intergroup conflicts in personal adjustment, the pathological consequences of restraint versus expression of hostile impulses and, more generally, the necessity for a culture to provide socially approved outlets for aggressive expression.

SUMMARY

In order to provide a model to account for the cathartic effects of an aggressive act and the conditions which result in a reduction in aggressive drive, the terms act of aggression and aggressive drive are analyzed. A definition of accidental and intentional aggression is offered. The latter category includes aggressive drive-mediated behavior and also instrumental aggression which denotes aggressive acts which are reinforced by the attainment of nonaggressive goals. It is argued that even in instrumental aggression the noxious consequences of the aggressive act are significant determinants of the aggressor's behavior because of the mediating function of the labeling of these consequences and because of the inhibitory effects of empathic responses. In defining aggressive drive, a distinction is made between expressive aggression and hostile aggression which is illustrated in the difference between the desire to hit and the desire to hurt. The essential element in expressive aggression is the topography of the response while in hostile aggression, the critical factor is the noxious consequences of the act.

Hostile aggression is assumed to be a learned drive whose primary but not sole antecedents are past exposure to punishment and present threats to self-esteem.

A stimulus-mediating response model of aggressive drive suggests a number of procedures for reducing aggression other than performance of an aggressive response. These include modification of the initial eliciting stimulus conditions and facilitation of responses which are incompatible with anger and hostility. The consequences of aggressive actions are more complex inasmuch as they may entail changes in the strength of the expressive impulse, in hostile aggressive drive, in aggression anxiety and in the habit strength of the aggressive response. Since different measures of aggression may be differentially responsive to these various processes and may yield diametrically opposite effects, the use of diverse measures of aggression is required to evaluate the cathartic effects of a vicarious or direct aggressive response upon aggressive drive.

REFERENCES

BANDURA, A., & WALTERS, R. H. *Adolescent aggression.* New York: Ronald Press, 1959.

BANDURA, A., & WALTERS, R. H. Aggression. In, *Child psychology, sixty-second yearbook of the national society for the study of education.* Chicago: Univer. Chicago Press, 1963. Pp. 364–415.

BANDURA, A., ROSS, DOROTHEA, & ROSS, SHEILA A. Transmission of aggression through imitation of aggressive models. *J. abnorm. soc. Psychol.,* 1961, **63**, 575–582.

BANDURA, A., ROSS, DOROTHEA, & ROSS, SHEILA A. Imitation of film-mediated aggressive models. *J. abnorm. soc. Psychol.,* 1963, **66**, 3–11.

BERGER, S. M. Conditioning through vicarious instigation. *Psychol. Rev.,* 1962, **69**, 450–466.

BERKOWITZ, L., GREEN, J. A., MACCAULEY, J. R. Hostility catharsis as the reduction of emotional tension. *Psychiatry,* 1962, **25**, 23–31.

BERKOWITZ, L. *Aggression: A social psychological analysis.* New York: McGraw-Hill, 1962.

BERKOWITZ, L. Some factors affecting the reduction of overt hostility. *J. abnorm. soc. Psychol.,* 1960, **60**, 14–21.

BERKOWITZ, L. The expression and reduction of hostility. *Psychol. Bull.,* 1958, **55**, 257–283.

BUSS, A. H. *The psychology of aggression.* New York: Wiley, 1961.

DARWIN, C. *The expression of the emotions.* London: John Murray, 1873.

DOLLARD, J., DOOB, L., MILLER, N., MOWRER, O., & SEARS, R., FORD, C. S., HOVLAND, C. I., & SOLLENBERGER, R. T. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.,* 1955, **50**, 3–11.

FESHBACH, S. The catharsis hypothesis and some consequences of interaction with aggressive and neutral play objects. *J. Pers.,* 1956, **24**, 449–462.

FESHBACH, S. The stimulating versus cathartic effects of a vicarious aggressive activity. *J. abnorm. soc. Psychol.,* 1961, **63**, 381–385.

FESHBACH, S., & SINGER, R. The effects of fear arousal and suppression of fear upon social perception. *J. abnorm. soc. Psychol.,* 1957, **55**, 283–288.

FESHBACH, S. The effects of emotional restraint upon the projection of happiness. *J. Pers.,* 1963, **31**, 471–481.

FREUD, S., & BREUER, J. On the psychical mechanisms of hysterical phenomena. In S. Freud, *Collected papers.* Vol. 1. London: Hogarth Press, 1948. Pp. 24–41.

GOODENOUGH, FLORENCE L. *Anger in young children.* (Institute Child Welfare Monograph Series, No. 9.) Minneapolis: Univer. Minnesota Press, 1931.

HOKANSON, J. E., & BURGESS, M. The effects of three types of aggression on vascular processes. *J. abnorm. soc. Psychol.,* 1962, **64**, 446–449.

HOKANSON, J. E., & SHETLER, S. The effect of overt aggression on physiological arousal. *J. abnorm. soc. Psychol.,* 1961, **63**, 446–448.

IRWIN, F. W. An analysis of the concepts of discrimination and preference. *Amer. J. Psychol.,* 1958, **71**, 152–163.

Irwin, F. W. On desire, aversion, and the affective zero. *Psychol Rev.,* 1961, **68,** 293–300.

James, W. The moral equivalent of war. In, *Memoirs and studies.* London: Longmans, 1917.

Kaufmann, H., & Feshbach, S. Displaced aggression and its modification through exposure to antiaggressive communications. *J. abnorm. soc. Psychol.,* 1963, **67,** 79–83. (a)

Kaufmann, H., & Feshbach, S. The influence of antiaggressive communications upon the response to provocation. *J. Pers.,* 1963, **31,** 428–444. (b)

Kenny, D. T. An experimental test of the catharsis theory of aggression. Ann Arbor, Mich.: University Microfilms, 1953.

Lövaas, O. I. Effect of exposure to symbolic aggression on aggressive behavior. *Child Develpm.,* 1961, **32,** 37–44.

Magaziner, D., & Feshbach, S. The reduction of hostility without catharsis. Paper presented at meetings of Eastern Psychological Association, Philadelphia, April 1961.

Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: A test of the sensory-tonic theory. *J. Pers.,* 1953, **21,** 400–410.

Murphy, Lois B. *Social behavior and child personality.* New York: Columbia Univer. Press, 1937.

Rapaport, D. The conceptual model of psychoanalysis. In D. Krech & G. S. Klein (Eds.), *Theoretical models and personality theory.* Durham: Duke Univer. Press, 1952. Pp. 56–81.

Rosenbaum, M. E., & DeCharms, R. Direct and vicarious reduction of hostility. *J. abnorm. soc. Psychol.,* 1960, **60,** 105–111.

Schaar, J. H. *Escape from authority.* New York: Basic Books, 1961.

Scott, J. P. *Aggression.* Chicago: Univer. Chicago Press, 1958.

Sears, R. R., Maccoby, Eleanor E., & Levin, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

Sears, R. R., Whiting, J. W. M., Nowlis, V., & Sears, Pauline S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.,* 1953, **47,** 135–234.

Sears, R. R. Personality development in the family. In J. M. Seidman (Ed.), *The child.* New York: Rinehart, 1958. Pp. 117–137.

Siegal, Alberta E. Film-mediated fantasy aggression and strength of aggressive drive. *Child Develpm.,* 1956, **27,** 365–378.

Singer, J. L., & Rowe, R. An experimental study of some relationships between daydreaming and anxiety. *J. consult. Psychol.,* 1962, **26,** 446–454.

Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following motor inhibitions and hyperactivity. *J. consult. Psychol.,* 1952, **16,** 359–364.

Stratton, G. M. *Anger: Its religious and moral significance.* New York: Macmillan, 1923.

Thibaut, J. W., & Coules, J. The role of communication in the reduction of interpersonal hostility. *J. abnorm. soc. Psychol.,* 1952, **47,** 770–777.

Werner, H., & Wapner, S. Sensory-tonic field theory of perception. *Psychol. Rev.,* 1952, **59,** 324–333.

Worchel, P. Hostility: Theory and experimental investigation. In D. Wilner (Ed.), *Decisions, values and groups.* Vol. 1. New York: Pergamon Press, 1960. Pp. 254–266.

Worchel, P. Status restoration and the reduction of hostility. *J. abnorm. soc. Psychol.,* 1961, **63,** 443–445.

(Received May 20, 1963)

Psychological Review
1964, Vol. 71, No. 4, 273–290

PERCEPTION, DRIVE, AND BEHAVIOR THEORY[1]

ROBERT B. ZAJONC
The University of Michigan

AND

DONALD D. DORFMAN
San Diego State College

The role of perceptual variables in Behavior Theory was examined in terms of their implications for stimulus-intensity generalization and discrimination learning. Examination of the so-called sensory-interaction effects seriously questioned the validity of the afferent stimulus-interaction postulate. The postulate predicts a greater generalization fall for reaction potential to a test stimulus which differs from CS by x units on one dimension and by y units on another than to a test stimulus which differs from CS by x units on the first dimension and is identical with respect to the other dimension. Considerations from the sensory-interaction literature suggest that under some conditions the opposite is true. Results of experiments on stimulus-intensity generalization and discrimination learning were found to conflict with the predictions derived from Behavior Theory, supporting the view based on sensory-interaction effects. Treating drive (D) in terms of stimulus component (S_D) and suggesting that its effects might under some circumstances be a special case of a sensory-interaction effect, the equivalence of the consequences of stimulus intensity and drive intensity were examined. Experimental results supported the view proposed. An argument was advanced for an increased role of perceptual variables in Behavior Theory.

In his last book Hull (1952) wrote:

> Perhaps one of the reasons why the failure to distinguish sharply between stimulation and perception does not interfere any more markedly with the validity of the present deductions is that the elements of both stimulation and generalization are explicitly included in the system [p. 354].

The distinction between perception and stimulation has, in fact, been regarded as unnecessary in behavior theory for many years. There is, however, a growing body of literature, never directly associated with behavior theory, which might question this position. This body of literature, generally referred to under the rubric of sensory interaction, contains a fair amount of evidence that perceived magnitudes of stimuli impinging on a given sense modality are often markedly influenced by the simultaneous (or near-simultaneous) stimulation of another sense modality. The causes of sensory interaction effects are far from well understood, and the effects are by no means universal or unequivocal. But one can hardly question the fact that sensory interaction effects do occur.[2] Viewed in the light of the Hullian position the results of sensory interaction research have potentially significant implications for behavior theory. The present paper attempts to examine some of these implications and to demonstrate that the failure to distinguish between stimulation and

---

[1] Studies reported here were part of a research program supported jointly by Grants AF-49(638)367 and G-4951 from the Air Force Office of Scientific Research and the National Science Foundation, issued to the Research Center for Group Dynamics.

[2] The reader can easily observe a sensory interaction effect some quiet evening by lifting his telephone receiver, listening to the dial tone, fixating at a light bulb, and having someone switching it on and off in rapid succession. The amplitude of the dial tone will appear to oscillate in phase with the light. For best results no other light source should be present.

perception does in fact interfere with the deductions of behavior theory.

Sensory Interaction and Stimulus Generalization

The sensory interaction literature has a direct relevance for Postulate XI: Afferent stimulus interaction (Hull, 1952, p. 11). Reviews and summaries of this literature (Carter, 1962; Dufy, 1962; Gilbert, 1941; London, 1954; Ryan, 1940) seem to indicate that in the majority of instances the simultaneous or near-simultaneous extraneous stimulation results in the lowering of sensory thresholds. Also, when subjects are required to estimate stimulus magnitudes presented at supraliminal levels, simultaneous extraneous stimulation results in the overestimation of these magnitudes. In a small number of cases one also finds absolute threshold increase or magnitude underestimation. About a third of the experiments show no effects.

While these threshold and estimation effects are probably manifestations of the same perceptual consequences of the extraneous stimulation, in that the obtained threshold changes reflect changes in the perceived or "effective" stimulation, rather than improvement or deterioration of the sensory capacity,[3] it is the stimulus magnitude over- and underestimation that has a more direct bearing upon the afferent stimulus interaction postulate. The postulate reads as follows:

> All afferent impulses (s's) active at any given instant, mutually interact converting each other into s's which differ qualitatively from the original $\breve{s}$'s so that a reaction potential (${}_sE_R$) set up on the basis of some afferent impulse (s) will show a generalization fall to ${}_{\breve{s}}E_R$ when the reaction (R) is evoked by the other afferent impulse ($\breve{s}$) [Hull, 1952, p. 11].[4]

While the postulate assumes a *qualitative* change in the molar impulse which arises when the original stimulus is presented together with previously neutral stimuli, it certainly must, as a special case, subsume *quantitative* changes. For in the last analysis a qualitative change is nothing but a cluster of quantitative changes too complex to be conveniently described in quantitative terms. Allowing the postulate to deal with quantitative changes in the molar impulse will enable us on the basis of sensory interaction literature to derive some new implications.

Consider a given sensory interaction effect on stimuli lying on a particular intensity continuum, A. Stimulus intensities on this continuum are known to be overestimated when judged in the presence of some extraneous stimulation, B. Low magnitudes of the extraneous stimulation,

---

[3] The effects of factors such as extraneous stimulation or drive on absolute threshold and magnitude judgments are subject of a separate problem. It might be pointed out in passing, however, that when we observe a lowering of absolute threshold which obtains as a consequence of some manipulation, such as increased drive level, for example, the results cannot always be interpreted as an "improvement" in perception. The same manipulation introduced when the subject is making magnitude judgments of these stimuli, supraliminally presented, may result in the overestimation of these magnitudes. Overestimation can hardly be interpreted as an "improvement," since it represents a decrement in accuracy.

[4] As paraphrased by Hilgard (1956) the afferent stimulus interaction postulate reads more clearly:

> When a response has been conditioned to a stimulus (S_1) and one or more previously neutral stimuli (S_2, S_3, $\cdots$) are presented along with S_1, the afferent impulses from this combination of stimuli interact to yield a new molar impulse (s). This new impulse is equivalent to a stimulus at a greater or less distance from S_1 on a qualitative continuum. The resulting generalized reaction potential to s will be smaller than to S_1, depending upon their remoteness from each other.

B_1, result in minimal overestimation, and high magnitudes, B_2, result in a larger overestimation of the A stimuli. Suppose further that a response is conditioned to a particular A stimulus when the latter is presented together with B_1. Following training, this A stimulus is again presented but now together with B_2, and performance is observed. Since the perceived or "effective" intensity of the A stimuli is known to increase the presence of B_2, performance to the particular CS will show a generalization fall when the extraneous stimulus intensity is increased. Thus far the prediction is entirely in agreement with the afferent stimulus interaction postulate. However, since we are dealing with presumably known quantitative changes in the "effective" stimulation, we are also in a position of making specific predictions about generalized responses to A stimuli other than the particular CS; for instance, $\text{CS} - a$, $\text{CS} - 2a$, $\text{CS} - 3a$, $\cdots$, $\text{CS} + a$, $\text{CS} + 2a$, $\text{CS} + 3a$, $\cdots$, where a is a magnitude on the A dimension, expressed perhaps in log units to satisfy Postulate X (Hull, 1952, p. 11).

The generalization fall in performance to the CS had been obtained because its perceived intensity increased when it was presented together with B_2. But other A stimuli will also be perceived as more intense in the presence of B_2 than in the presence of B_1. Since $\text{CS}-a$, $\text{CS}-2a$, $\cdots$, would in the presence of B_2 appear more intense, they would become effectively *more similar* to the trace of the CS maintained from training, where it was presented in the company of the weaker extraneous stimulation, B_1. Let us, for the purposes of argument, assume that the psychophysical change in the perception of the A stimuli occurring when B_1 is replaced by B_2 is equivalent to one a unit, and that such changes are more or less uniform for the range of values employed in the generalization test, such that any given stimulus intensity on the A continuum, S_{A_i}, when presented with B_1 is perceptually equivalent to $S_{A_i} + a$, when B_1 is replaced by B_2. Thus, generalized responses to $\text{CS} - a$, $\text{CS} - 2a$, $\cdots$, will be more probable (or stronger) when obtained in the presence of B_2 than when obtained in the presence of B_1. Responses to $\text{CS} - a$ obtained in the presence of B_2 might be as probable as responses to the CS itself when the latter is presented in the company of B_1. Responses to $\text{CS}-2a$ obtained in the presence of B_2 will be as probable as those obtained to $\text{CS} - a$ in the presence of B_1, etc.

Since $\text{CS} + a$, $\text{CS} + 2a$, $\cdots$, would in the presence of B_2 also increase in their perceived intensity by one a unit, they would becomes less similar to the trace of the original CS, and the generalized responses elicited by them would be weaker than when observed in the presence of B_1. Consequently, for the stimulus continuum employed in the test, we obtain not a fall in the generalization gradient but a lateral displacement of the gradient toward the lesser stimulus intensities of the A continuum, with mode at $\text{CS} - a$.

These predictions conflict with ones derived from Postulate XI. Consider only two stimuli, the CS and $\text{CS} - a$, which for the purpose of simplicity we will now refer to as S_1 and S_2, retaining the relation $S_1 > S_2$. Since the original training occurs in the presence of B_1, the trained response is attached to the compound S_1B_1. Above we predicted that the generalized reactions to stimuli *weaker* than the CS will be more probable when obtained under increased extraneous stimulation than when obtained under constant extraneous stimulation, or in Hullian terms that ${}_{S_2B_2}\underline{E}_R > {}_{S_2B_1}\underline{E}_R$. Essentially then we are predicting that in the case of a two-dimensional stimulus generalization, a test stimulus which differs from the CS on one dimension will produce a lower reaction potential than a test stimulus which differs from the CS to an equal extent on that dimension, but which also differs from the CS on one other dimension!

But we are dealing here with stimulus-intensity continua and therefore considerations deriving from stimulus-intensity dynamism (V) may complicate matters. A more critical examination of the Hullian prediction is therefore called for. If drive (D) and incentive (K) are held constant, we have the generalized reaction potential to S_2, obtained when extraneous stimulation remains the same from training to test, ${}_{S_2B_1}\underline{E}_R = {}_{S_1B_1}E_R \times 10^{-jd_S} \times V_{2B_1}$, where j is an empirical generalization constant, d_S the difference between S_1 and S_2 in log units, and V_{2B_1} the stimulus-intensity dynamism of the combined S_2 and B_1 stimulation. The generalized reaction potential to the same test stimulus but obtained under increased extraneous stimulation can be written as follows: ${}_{S_2B_2}\underline{E}_R = {}_{S_2B_1}\underline{E}_R \times 10^{-id_B} \times V_{2B_2}$, where i is another empirical generalization constant, d_B the difference between B_1 and B_2 in log units, and V_{2B_2} the stimulus-intensity dynamism of the combined S_2 and B_2 stimulation. To obtain the prediction based on sensory interaction effects from Hullian theory we must prove that ${}_{S_2B_2}\underline{E}_R > {}_{S_2B_1}\underline{E}_R$, or ${}_{S_2B_1}\underline{E}_R \times 10^{-id_B} \times V_{2B_2} > {}_{S_2B_1}\underline{E}_R$. But since both 10^{-id_B} and V_{2B_2} are equal to or smaller than unity, the above relation is false according to Hull's theory.

When a response is trained to an *A* stimulus in the presence of the stronger extraneous stimulation, B_2, and generalization tests are performed in the presence of the weaker extraneous stimulation, sensory interaction effects suggest that a mirror image of the former prediction will be obtained. Now the generalization gradient will show a lateral displacement toward the greater stimulus intensities with the peak of the gradient at CS + a. Under these conditions the responses to stimuli stronger than the CS will increase in probability when extraneous stimulation is decreased during generalization testing. Here the Hullian model is even less likely to predict an increment in generalized reaction potentials as a consequence of the decrement in extraneous stimulation. Now the difference between the CS and the testing stimulus as well as the reduction in the total stimulus-intensity dynamism both act to reduce ${}_{S}E_{R}$.

In an experiment whose attempt was to examine the above predictions (Dorfman, 1961a), the training CS and the generalization stimuli were slight electric shocks (from −.38 to −.02 log milliampere) of 1-second duration. The extraneous stimulation was a 45-decibel tone for B_1 and a 96-decibel tone for B_2, both at 1,000 cps, occurring also for a 1-second duration and simultaneously with the electric shocks. The distance between the stimuli selected for the generalization test (−.06 log milliampere) was established on the basis of a previous psychological investigation (Dorfman, 1961b) in which there were indications that a change from B_1 to B_2 resulted in an overestimation of shock intensities of approximately that order.

One group of subjects were trained under low extraneous stimulation, and another group under high extraneous stimulation. In order to avoid the problem of asymmetrical generalization gradients a procedure developed by Brown, Clarke, and Stein (1958) was employed. One verbal response (WIN) was reinforced to the −.20 log milliampere shock, and another (LOSE) to the shock of the lowest (−.38 log milliampere) and to the shock of the highest intensity (−.02 log milliampere).[5] The three training stimuli were presented one at a time in 15-second intervals, and simultaneously with either B_1 or B_2, depending on the experimental condition. In both groups subjects were given 12 training trials to the middle stimulus, and 7 training trials to each of the two extreme stimuli. The subjects were reinforced for correct responses by the experimenter saying "right" and for incorrect responses by the experimenter's saying "wrong."[6] In the generalization test each of the seven stimuli was presented six times: three times with B_1 and three times with B_2 in a random sequence. During test no reinforcement was given, however, interspersed among the testing trials were six reconditioning trials. The data for this experiment are reproduced in Figure 1, where it is clear that a lateral displacement of the gradients occurred when the extraneous stimulation was changed from training to test. When subjects are trained in the presence of low extraneous stimulation the modes and the medians of their generalization gradients are at significantly lower shock intensities for tests taken in the presence of strong extraneous stimulation than for tests taken in the presence of weak extraneous stimulation. Conversely, when training occurs under high extraneous stimulation, and when the test is given in the presence of the weaker tone, the modes and medians of the gradients are displaced toward higher shock intensities. When extraneous stimulation is increased generalized reactions to stimuli weaker than the CS gain in probability of emission, and when it is decreased reactions to stimuli stronger than the CS gain in probability. These effects were uniform throughout the course of the generalization tests. For purposes of comparison, the Hullian predictions for this − + − training procedure were computed, assuming the usual constants for generalization and for V (Hull, 1952), and using a formula suggested by Hull (1951, p. 94) for computing the combined V, involving both the S and the B intensities. These predictions are shown alongside the empirical results, and it is clear that the results are in disagreement with the implications of the Hullian model.

Drive and Stimulus Generalization

There has always been some ambiguity in predicting the effects of drive shift upon generalization. The energizing component of drive (D) leads to one set of predictions and the stimulus component of drive (S_D) to another. Because of the multiplicative laws, $E=D\times H$ and $E=D\times\bar{H}$, one would expect increments in generalized responses and in the generalization slope when drive is increased following training, and decrements when drive is decreased. No displacement of the modal values of the generalization gradients would be expected. In support of this prediction Spence (1958) reports an experiment by Newman (1955) who ob-

---

[5] This experiment was also performed using a conventional generalization procedure involving only one response, and essentially the same results were obtained (Karp, 1961).

[6] There were fewer than 10% of incorrect responses during the initial training, the discrimination performance reaching 100% accuracy toward the end of the training.

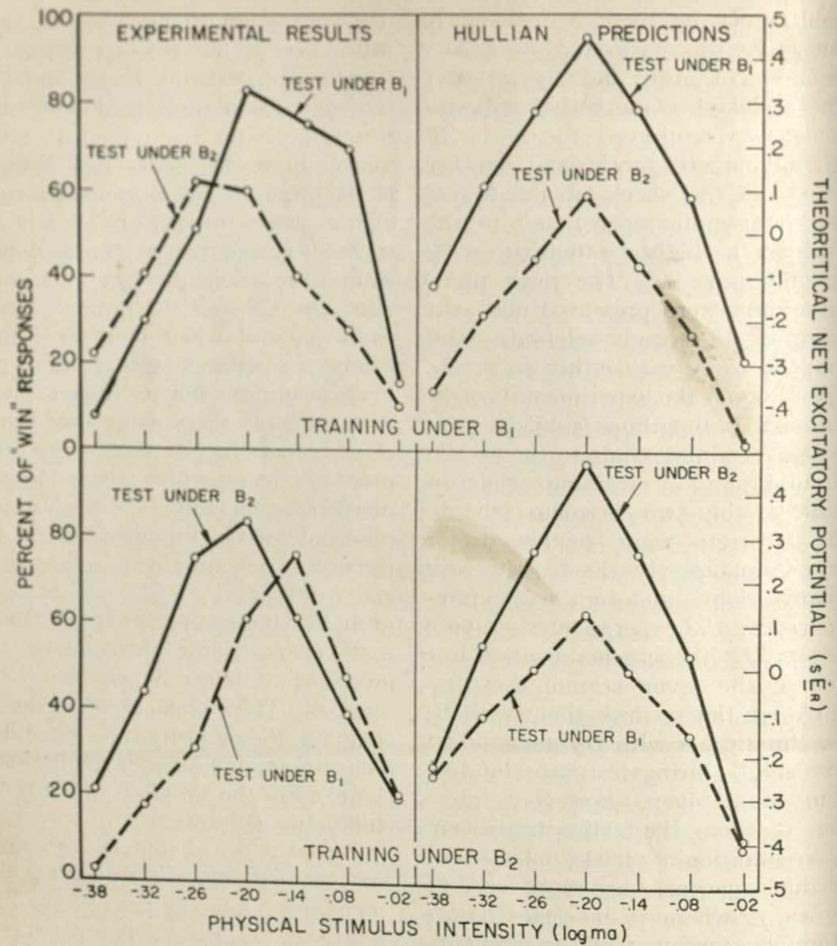


FIG. 1. The effect of change in extraneous stimulation (loudness) on generalization along the shock dimension: Experimental results and theoretical predictions from Hull.

served generalization along the size dimension for rats whose hunger level was either increased or decreased following training.

From the point of view of the drive stimulus, S_D, however, one would expect a decrement in response strength regardless of whether drive is increased or decreased following training. Since increments as well as decrements in drive represent changes in the molar impulse, the law of afferent stimulus interaction applies, predicting that any change in the molar impulse produces a generalization fall. It seems that Hull must have given more weight to the role of S_D, for in Corollary XI he states:

> When a habit is set up in association with a given drive intensity and its strength is tested under a different drive intensity, there will result a falling gradient of $_s\bar{H}_R$ and $_s\underline{E}_R$ [Hull, 1951, p. 90].

Apparently this corollary was established solely on the basis of empirical results such as those of Yamaguchi (1952), for one aspect of the corollary seems to contradict the formal model. The corollary is consistent with the model with respect to the prediction of a falling ${}_{S}\bar{H}_{R}$ gradient. However, it is difficult to see that the model can also predict a falling ${}_{S}\underline{E}_{R}$ gradient. Consider a reaction potential set up under a given drive level, D_1, and a subsequent test under an increased drive, D_2. The corollary predicts that ${}_{ss_{D_1}}E_R > {}_{ss_{D_2}}\underline{E}_R$, where S_{D_1} and S_{D_2} are the two drive stimuli. Expanding we have: ${}_{ss_{D_1}}H_R \times D_1 \times V_1 > {}_{ss_{D_1}}H_R \times 10^{-bd_{S_D}} \times D_2 \times V_1$, where b is an empirical generalization constant, and d_{S_D} the difference between the drive stimuli. The above relation simplifies to $D_1 > 10^{-bd_{S_D}} \times D_2$. Fixing D_1 at 1.0 and assuming that S_D is equal to $\log D$, we obtain the result that the empirical generalization constant b must be smaller than -1.0 for the relation to hold. Since the empirical results on generalization consistently indicate that the constant is seldom smaller than $-.15$, the corollary cannot be justified. Moreover, the corollary conflicts with the data reported by Newman (1955). On the other hand, working with the stimulus sampling model and treating drive as a set of irrelevant stimulus elements, Estes (1958) arrived at a prediction for generalization gradients obtained after a drive shift. Contrary to the above inferences, increments in drive following training were predicted by him to result in a shallower gradient with response probability somewhat depressed at the CS, while decrements in drive in a steeper gradient, also with a depressed response probability at the CS.

If, in the case of drive shift, the most significant aspect of drive is, as Estes argues, its stimulus element, S_D may be viewed simply as an extraneous stimulus such as we dealt with in the above experiment by Dorfman (1961a). If as such S_D leads to the overestimation of other stimulus magnitudes, then an increment in drive following training should result in a lateral displacement of the generalization gradient in the direction of lower stimulus magnitudes, and a decrement in drive in the direction of higher stimulus magnitudes. We would thus expect essentially the same effects whether we manipulate some irrelevant extraneous stimulation of irrelevant drive. It should be noted that neither the inference based on S_D effects, nor one based on the multiplicative law predict a lateral displacement of the generalization gradient following a drive shift.

In the present program of research some experiments dealt with this problem (Carter & Zajonc, 1962; Matlin, 1960; Platz, 1962; Platz & Zajonc, 1961; Zajonc & Cross, 1964). In one of the experiments (Platz, 1962) subjects were first given differential fear conditioning in a series of 29 trials in which one of two lights was followed by a fairly strong electric shock. The duration of the light was .5 second, and its separation from the shock offset varied randomly from zero to five seconds for each subject. The shock intensity was set according to the tolerance of each subject, found prior to the fear conditioning. Measures of GSR were taken to assess the relative effectiveness of the differential fear conditioning. Following this fear conditioning essentially the same procedure as that in Dorfman's experiment (1961a) was employed. A particular verbal response (HIT) was reinforced over a series of 15 trials to a 45-decibel tone and another verbal response (MISS) was reinforced to a 42.5-decibel and to a 50.0-decibel

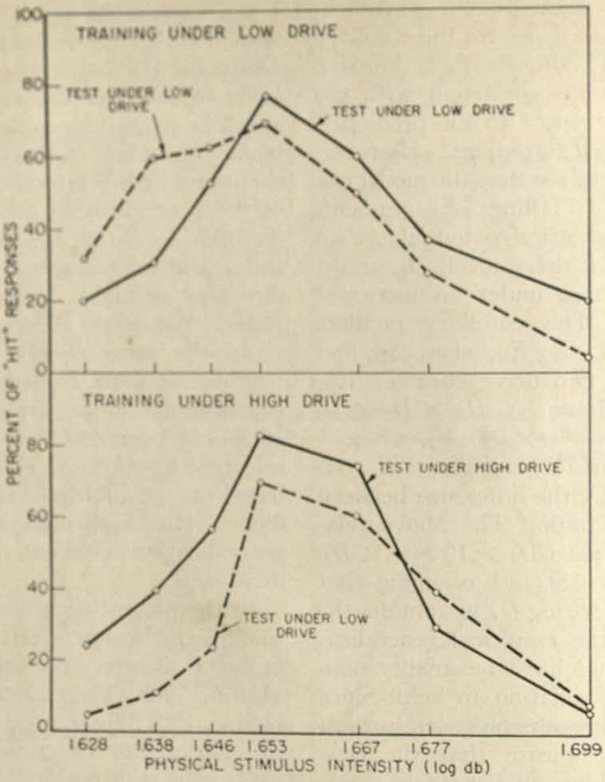


FIG. 2. The effect of change in drive (fear) on generalization along the loudness dimension.

tone over a series of 12 trials each. In one group of subjects the middle and the extreme stimuli were presented together with the previously shocked light (with the shock now absent), and in another group the three training stimuli were presented with the previously nonshocked light. As in Dorfman's experiment, following the − + − training Platz observed the proportion of HIT responses to seven test stimuli ranging from 42.5 to 50.0 decibels. All auditory stimuli were given at 1,000 cps. The test stimuli were given in a random order in 15-second intervals. The shock (or nonshock) light was of .5-second duration, and it preceded the tone stimulus by one second. On half of the trials the test stimuli were presented together with the shocked light (but without the shock), and on the other half in the presence of the nonshocked light. Interspersed among the generalization trials were also some reconditioning trials to maintain the differential fear reaction to the two lights. On each generalization trial GSR measures were taken.

Figure 2 shows the results of the drive shift upon the proportion of generalized HIT responses. A lateral displacement similar to that found when extraneous stimulation was manipulated is observed. When drive is increased from training to test, the curve is displaced toward the weaker tones. When drive is decreased, however, the gradients show a displacement in the direction of the louder tones. But Platz's results, although at an acceptable level of significance,[7] seem to be less pronounced than those of Dorfman. When drive is manipulated the peaks of the gradients did not shift as was the case when extraneous stimulation was manipulated. There is, however, a basic difference between the two

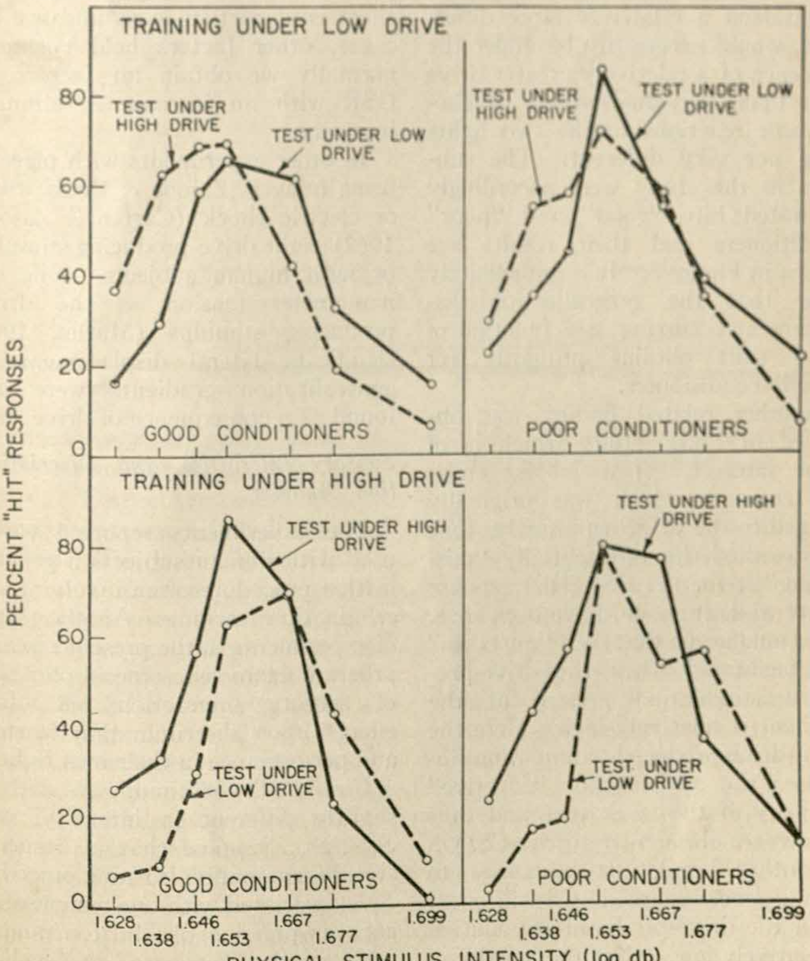


Fig. 3. The effect of change in drive (fear) on generalization along the loudness dimension for "good" and "poor" conditioners.

---

[7] The average median of the gradient of the group trained under high drive was displaced in the direction of higher stimulus intensities by .68 decibel when the test stimuli were given with the nonshock light, while the median of the group trained in the nonshocked light was displaced in the direction of lower intensities by .77 decibel when the stimuli were given in the presence of the shocked light. This effect was significant at .001 level ($F = 36.85, df = 1/38$).

manipulations. In Dorfman's experiment the difference in the levels of extraneous stimulation were constant for all subjects. In the experiment manipulating drive, the differences in conditioned fear depended on the success of the previous differential fear conditioning and its extinction status during the training and testing of responses to the auditory stimuli. Individuals who conditioned well, in that their GSRs to the shocked and nonshocked lights reached and maintained a relatively large difference, would necessarily be under the influence of a relatively greater drive shift than individuals whose psychogalvanic reactions to the two lights were not very different. The subjects in the study were accordingly separated into "good" and "poor" conditioners and their results are shown in Figure 3. It is immediately clear that the generalization displacement occurring as a function of drive shift obtains primarily for "good" conditioners.

Another related finding was obtained in Platz' study which is of some interest. If a subject emits the response which was originally trained to the CS when a *weaker* tone is presented, he is essentially "mistaking" it for the CS. His response is HIT while it should have been MISS. One could argue that such "mistakes" are facilitated when the drive-produced stimulation present at the moment is relatively high. Here the high drive-produced stimulation increases the apparent or "effective" intensity of the weak tone and thus makes it more similar to the CS. On the other hand, HIT responses to stimuli *louder* than the CS will occur when the drive-produced stimulation is relatively low. Here the low drive-produced stimulation decreases the apparent intensity of the test stimulus and thereby makes it similar to the CS. To the extent that the GSR measures the intensity of this drive-produced stimulation, we would expect that the GSR obtained when the subject emits a HIT response should *decrease* with the intensity of the test stimulus. The level of GSR observed when HIT responses were given by subjects to the seven auditory stimuli are shown in Figure 4. The relation expected from the sensory interaction point of view is clearly evident. This effect is of particular significance because, other factors held constant, normally we obtain an *increase* in GSR with an increase in stimulus intensity.

In other experiments with pigeons, using hunger (Zajonc & Cross, 1964) or electric shock (Carter & Zajonc, 1962) as the drive-producing stimulus, or with human subjects, using dynamometer tension as the drive-producing stimulus (Matlin, 1960) significant lateral displacements of generalization gradients were also found as a consequence of drive shift.

Sensory Interaction and Discrimination Behavior

The experiments reported earlier used with human subjects a generalization procedure often involving discrimination learning. Another series of experiments in the present research program examined some implications of sensory interaction and drive effects upon discrimination learning and performance in their own right.

Consider two stimuli, S_1 and S_2, slightly different in intensity, with $S_1 < S_2$. Suppose that in a successive discrimination learning procedure S_1 is associated with one response, R_1, such as pressing one button, and S_2 with another response, R_2, such as pressing another button. Suppose further that S_1 and S_2 are alwa

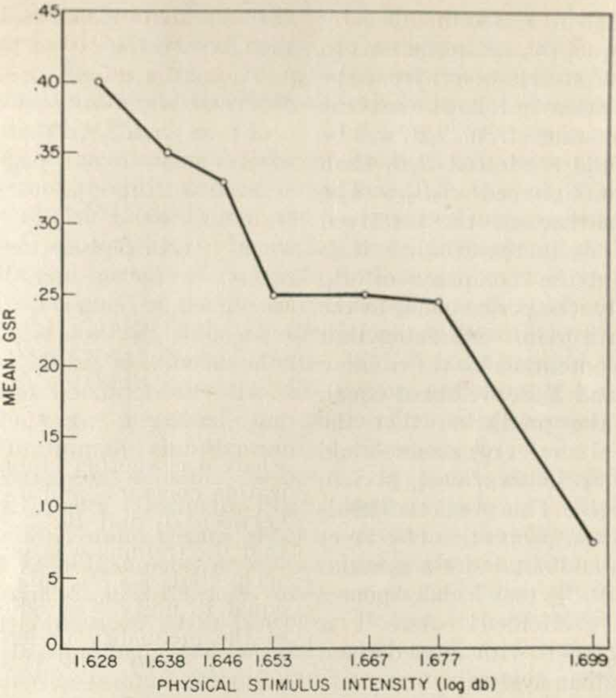


FIG. 4. Mean GSR (in micromhos) accompanying "hit" responses as a function of stimulus intensity.

presented together with some extraneous stimulation B_1, low in intensity. Correct responses to the stimuli are rewarded and incorrect punished. The pairs S_1B_1 and S_2B_1 occur in a random sequence and equally often over a series of trials. Interspersed among these trials, however, are also trials on which the two stimuli, S_1 and S_2, one at a time, are presented simultaneously with extraneous stimulation, B_2, considerably more intense than B_1. These trials occur in a random sequence and as often as those accompanied by B_1. Whenever S_1B_2 or S_2B_2 are presented neither positive nor negative reinforcement is administered, although the subjects are still required to select one of the two responses. We are concerned with the prediction of discrimination performance with respect to the nonreinforced pairs, S_1B_2 and S_2B_2.

Since we are dealing with discrimination learning, the critical element of the situation is the distance between S_1 and S_2. It was noted above that sensory interaction effects lead to changes in the perceived intensities of stimuli, and we must therefore consider not the physical but the "effective" or the perceived distance between stimuli. The assumption that extraneous stimulation increases the perceived intensities of S_1 and S_2 leads us to predict that the "effective" distance from S_2B_1 will be greater for the reinforced S_1B_1 than for the nonreinforced S_1B_2. The change in the "effective" distance obtains because the perceived intensity of S_1

increases when it is accompanied by B_2. Consequently, performance to S_1B_2 will suffer. However, from the same assumption it follows that the "effective" distance from S_1B_1 will be smaller for the reinforced S_2B_1 than for the nonreinforced S_2B_2. The effect obtains because the perceived intensity of S_2 increases when it is presented in the company of B_2. Consequently, the performance to the pair S_2B_2 will gain. Assuming that the reaction potentials for the training pairs S_1B_1 and S_2B_1 are about equal, we obtain the prediction that the proportions of correct responses should be ordered as follows: $S_2B_2 > S_2B_1 \approx S_1B_1 > S_1B_2$. This prediction holds that the highest proportion of correct responses should obtain for a stimulus compound, S_2B_2, to which responses were never reinforced! And it is again in conflict with one derived from the Hullian system.

It suffices to show that at least one of the inequalities of the above relation cannot be derived from Hull's postulates. Let us examine the ability of the Hullian model in predicting ${}_{S_2B_1}\dot{E}_{R_2} < {}_{S_2B_2}\dot{E}_{R_2}$, where ${}_S\dot{E}_R$ is the net discriminatory reaction potential. In terms of the Hulliam law of discrimination learining (Hull, 1952, pp. 85–92) we have the net discriminatory reaction potential for S_2B_1:

$${}_{S_2B_1}\dot{E}_{R_2} = {}_{S_2B_1}E_{R_2} \dot{-} {}_{S_2B_1}E_{R_1}.$$

Assuming that the product $D \times K$ is constant:

$$\begin{aligned}{}_{S_2B_1}\dot{E}_{R_2} = {}_{S_2B_1}H_{R_2}V_{2B_1} \dot{-} {}_{S_1B_1}H_{R_1}&\\ \times V_{1B_1} \times V_{2B_1} \times 10^{-bd_S}&,\end{aligned}$$

where the first subscript for V refers to the discriminanda S_1 and S_2, and the second to the extraneous stimuli, B_1 and B_2 (V being the stimulus-intensity dynamism of the combined S and B stimulation), b is an empirical generalization constant, and d_S the distance between S_1 and S_2. We also have:

$$\begin{aligned}{}_{S_2B_2}\dot{E}_{R_2} &= {}_{S_2B_2}E_{R_2} \dot{-} {}_{S_2B_1}E_{R_1},\\ &= {}_{S_2B_2}\bar{H}_{R_2} \times V_{2B_2}\\ &\qquad \dot{-} {}_{S_2B_2}\bar{H}_{R_1} \times V_{2B_2},\\ &= {}_{S_2B_1}H_{R_2} \times V_{2B_1} \times 10^{-ad_B}\\ &\quad \times V_{2B_2} \dot{-} {}_{S_2B_1}\bar{H}_{R_1} \times V_{2B_1}\\ &\qquad \times 10^{-ad_B} \times V_{2B_2},\\ &= {}_{S_2B_1}H_{R_2} \times V_{2B_1} \times 10^{-ad_B}\\ &\quad \times V_{2B_2} \dot{-} {}_{S_1B_1}H_{R_1} \times V_{1B_1}\\ &\quad \times 10^{-bd_S} \times V_{2B_1}\\ &\qquad \times 10^{-ad_B} \times V_{2B_2},\end{aligned}$$

where a is another empirical generalization constant, and d_B the distance between B_1 and B_2. Let us assume that the training was sufficiently long such that ${}_{S_1B_1}H_{R_1}$ and ${}_{S_2B_1}H_{R_2}$ are equal to unity. Dividing both sides of the inequality under examination by $V_{2B_1} \times 10^{-ad_B} \times V_{2B_2}$ we have:

$$\frac{1 \dot{-} V_{1B_1} \times 10^{-bd_S}}{V_{2B_2} \times 10^{-ad_B}} < 1 \dot{-} V_{1B_1} \times 10^{-bd_S}.$$

Since the denominator of the left side of the inequality must be smaller than unity, the inequality is false.

Consider further a situation where the reinforcement is administered only in the presence of the stronger extraneous stimulation, and we are concerned with the prediction of performance to the nonreinforced pairs, S_1B_1 and S_2B_1. Since a decrease in extraneous stimulation results in the "effective" weakening of S_1 and S_2 when they are presented with B_1, the distance between the pairs S_1B_2 and S_2B_2 will be smaller than between S_1B_1 and S_2B_2, and greater than between the pairs S_1B_2 and S_2B_1. Here then we predict that the highest percentage of correct responses will obtain for the nonreinforced pair S_1B_1. Thus, when reinforcement is given only in the presence of B_2, the net discriminatory reaction potentials are ordered as follows: ${}_{S_1B_1}\dot{E}_{R_1} > {}_{S_1B_2}\dot{E}_{R_1} \approx {}_{S_2B_2}\dot{E}_{R_2} > {}_{S_2B_1}\dot{E}_{R_2}$. It can be shown, as above, that the Hullian model predicts ${}_{S_1B_1}\dot{E}_{R_1} < {}_{S_1B_2}\dot{E}_{R_1}$.

Several experiments using the above paradigm were recently carried out, and their results seem rather consistent with expectations. In each experiment one group of subjects were reinforced only in the presence of weak, and another only in the presence of strong extraneous stimulation. In all experiments the discriminanda were presented for a 1-second duration and simultaneously with the extraneous stimuli. In one of the experiments (Dorfman, Platz, & Zajonc, 1962), mild electric shocks (.92 and 1.32 milliampere) served as the discriminanda and tones of weak and strong intensities as extraneous stimuli. In another, two stimuli slightly differing in loudness (51 and 55 decibels, 1,000 cps) were used as the discriminanda and two levels of illumination provided extraneous stimulation. In a third experiment two similar levels of brightness of a patch of light were used as discriminanda, and two levels of auditory stimulation served as B_1 and B_2. On each trial one of the four S_iB_i combinations occurred, all equally often and in a random sequence over a series of 128 trials. Always, in one group of subjects reinforcement was given in the presence of B_1 only, and in another in the presence of B_2 only. The percentage of correct responses to the S_iB_i pairs found in these three experiments is summarized in the upper portions of Figure 5. The two first experiments clearly show the expected pattern of response probabilities. The results of the third experiment are somewhat weaker, although the same trend is still to be observed. In the second experiment (where auditory stimuli served as the discriminanda and illumination as the extraneous stimulation, and when reinforcement was administered under B_1) the difference in correct responses between the nonreinforced pairs S_1B_2 and S_2B_2 is over 30%, with correct responses to S_1B_2 being only slightly better than chance and those to S_2B_2 reaching almost 90%. It should be noted that the above response measures are based always on the entire set of trials, including early blocks when performance to all four S_iB_i combinations does not exceed the chance level.

It might be argued that in the present series of discrimination experiments the subjects may have used the irrelevant extraneous stimulation as an additional, albeit improper, cue. That is, for instance a subject who is only reinforced in the presence of the weaker stimulus, and is asked to choose between R_1 and R_2 when the pair S_1B_2 is presented, might when uncertain be tempted to select R_2 rather than the correct R_1, simply because the extraneous component of the pair is clearly strong, and because R_2 was always reinforced to the stronger of the two stimuli, namely S_2. Performance to S_1B_2 would therefore suffer. When presented with S_2B_2 and also uncertain, he might also choose R_2 in preference to R_1 for the same reason. This tendency, however, would benefit the performance to S_2B_2. The opposite tendencies may obtain when reinforcement is given only in the presence of B_2. Here performance to S_1B_1 would benefit and to S_2B_1 suffer. To examine these possibly confounding effects, the second of the previous discrimination experiments was repeated using the same two auditory stimuli as S_1 and S_2, and the same two levels of illumination as B_1 and

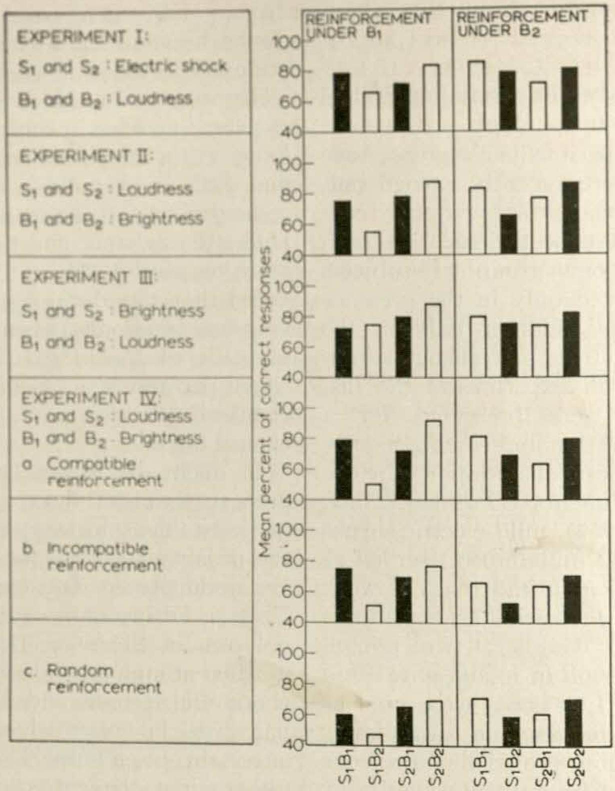


FIG. 5. Discrimination performance under extraneous stimulation in four experiments.

B_2, and the following preexperimental procedure. Before the experiment proper and before introducing S_1 and S_2 at all, the subjects were trained to respond to B_1 and B_2 in the absence of the discriminanda. In one group (Compatible Reinforcement) the response which was later to be associated with the weaker S_1 was reinforced to B_1 and the response to be later associated with the stronger S_2 to B_2. In another group (Incompatible Reinforcement) the response to be later associated with the weaker S_1 was reinforced to B_2, and that to be later associated with S_2 to B_1. In a third group (Random Reinforcement) half of the R_1 responses was randomly reinforced to B_1 and half to B_2, and similarly for the R_2 responses. This training was carried out for 60 trials. In the second part of the experiment a discrimination learning procedure identical to that in Discrimination Experiment II was used.

In all groups the effect found previously was observed. As can be seen in the lower portion of Figure 5, the pattern of differences between the proportions of correct responses is most pronounced for the Compatible Reinforcement group, showing that in the original experiments at least a part of the effect may be attributed to the tendency of responding on the basis of extraneous stimulation when a discrimination with respect to S_1 and S_2 is difficult. However, the presence of the differential response percentages to the four S_iB_i pairs, consistent with previous findings and found for all three groups, strengthens the conclusion that the effects previously obtained were in large part due to a sensory-interaction influence.

CONCLUSIONS

The results of the experiments presented above seem to argue for an increased role of perception in Behavior Theory. In particular they argue for an increased role of the afferent stimulus interaction postulate, although the postulate may have to be somewhat modified. Afferent impulses (s's) present at a given time do seem to interact and seem to convert one another. But it is also apparent that, at least with respect to the continua used in the above research, they convert one another into $\breve{s}$'s which are not qualitatively but quantitatively different from the original s's. Consequently, while the prediction of the fall in reaction potential (${}_s\underline{E}_R$) resulting when the response is elicited by $\breve{s}$ is correct, one cannot readily extend this prediction to the generalized reaction potential, ${}_s\underline{E}_R$. Depending upon the nature of the afferent interaction, the test stimulus, and the levels of extraneous stimulation present during training and testing, ${}_s\underline{E}_R$ may either show a fall or a rise. Exact predictions of ${}_sE_R$ and ${}_s\underline{E}_R$ would therefore require exact knowledge of the specific sensory-interaction effects which obtain for the particular eliciting and extraneous stimuli under consideration. Unfortunately, there are at the moment no systematic principles governing such sensory-interaction effects. For each pair of stimuli specific previous knowledge of interaction effects must be available.

The afferent stimulus interaction postulate holds specifically that ${}_{\breve{s}}E_R = {}_sE_R \times 10^{-jd}$, where j is an empirical generalization constant, and d the difference between s and $\breve{s}$ (Hull, 1951, p. 94). The only modification required is the addition of the qualification that at times a generalization gain obtains when responses are made to $\breve{s}$, and that d be considered in terms of "effective" or perceived rather than in terms of physical distance.

The increased significance of the afferent stimulus interaction postulate may have wider implications. Perceptual constructs and variables have never played important roles in Behavior Theory. Among the formal principles of Behavior Theory only the afferent-stimulus-interaction and the stimulus-intensity-dynamism postulates might be "perceptual" or "quasi-perceptual" in nature. Of these two, the former remained largely idle, while the latter was shown to be in conflict with experimental results (Nygaard, 1958), and ultimately unnecessary for the explanation of the data which prompted its introduction (Champion, 1962; Logan, 1954; Perkins, 1953). The idleness of the afferent stimulus-interaction postulate placed the eliciting stimulus in a rather enviable position among other variables. We began to view it as a stable and dependable component of the system. When experimental variables popular in behavioral research, and in particular drive, are manipulated, they are conventionally viewed as leaving the eliciting stimulus untouched. Rather, the principal attention is directed to how these variables affect the output side. Sensory-interaction data indicate that the same stimulus administered under two different background conditions is simply not the same stimulus. At least, it is not perceived or reacted to as if it were the same stimulus.

A further implication follows from the research presented in this paper. It is quite well established that the strength of learned responses is markedly influenced by drive (Spence, 1956). It is also rather well established that with drive level held constant, the intensity of the eliciting stimulus influences the strength of learned responses in very much the same way. Estes (1958) made this observation in a somewhat different context. The former effect has been commonly dealt with by means of the multiplicative law. The latter effect, on the other hand, has been originally treated in terms of stimulus-intensity dynamism and later by means of the discrimination principle elaborated by Perkins (1953) and Logan (1954). In the light of the results presented here, the communality of drive and stimulus-intensity effects need not remain simply accidental. The high rate of responding under high drive need not be entirely due to the energizing force of *D*. These effects might in some degree reflect the increased perceived intensity of the eliciting stimulus, which derives from its interaction with the drive-stimulus. Although the problem is more complex for acquisition, here too the possibility exists that stimulus intensity and drive can be "traded off" for one another *because* the latter increases the "effective" intensity of the eliciting stimulus. Thus, the effects of stimulus intensity and drive on acquisition and performance may prove to be equivalent not only in their consequences, but in part also in their immediate antecedents.

The possibility that the drive stimulus may produce sensory interaction effects is strengthened by recent neurophysiological evidence. On the one hand Steller (1954) argued that the state of drive of the organism is directly correlated with the amount of activity in the hypothalamus. Hebb (1955) and Lindsley (1957) also view general level of motivation as reflected in cortical arousal. On the other hand, many studies show perceptual effects due to neural activation. As a matter of fact, afferent stimulus interaction was at first presented under the label of afferent *neural* interaction. Fuster (1958), for instance, reports that discrimination is facilitated by a mild stimulation of the brain stem, and Boswell (1958) that recognition thresholds are lowered when obtained during the excitatory phase of the alpha rhythm. Thus, to the extent that drive, as well as extraneous stimulation, produce increased neural activity, and to the extent that such arousal interacts with other sensory inputs, drive, extraneous stimulation, and the eliciting stimulus intensity might all show equivalent effects.

The equivalent effects of drive and extraneous stimuli on stimulus reception and transmission may also have some implications for the work which, since the classical Bruner-Goodman experiment (1947) on the estimation of coin sizes, has concentrated on demonstrating motivational effects in perception. Although the role of motivational influences in perception is constantly being challenged by response-oriented interpretations, the argument presented here would certainly support the possibility of motivational effects in perception, such as accentuation of stimulus magnitudes and the lowering of sensory thresholds under increased motivation. But a departure from the New Look formulation is also indicated. Traditionally, the general hypothesis that motivational factors influence perception requires that in order for a motivational state to exert some effect upon the perception of a stimulus, the stimulus must in some relevant, perhaps instrumental, way be related to the manipulated drive. The above restriction, which is essential to the New Look interpretation of motivational effects in perception may be quite unnecessary. Any drive, provided its S_D interacts with the judged stimulus may exercise such effects as magnitude over- or underestimation and changes in recognition threshold. This latter thought is perhaps less enchanting from clinical point of view, but on the other hand, there is little reason to doubt it. Differential perceptual performance associated with personality differences need not be interpreted in terms of the differential psychological relation of the different personalities to the stimulus object. Rather, it can be viewed as a consequence of differences in arousal, drive, or sensitivity to sensory-interaction effects. It is still possible that when the judged stimulus has instrumental significance for the motivational state in question, some special effects might be obtained. It is more likely, however, that these effects would be limited to influences upon processes peripheral to what might be called "immaculate" perception, such as orienting, attending, and the like.

REFERENCES

BOSWELL, R. S. An investigation of the phase of the alpha rhythm in relation to visual recognition. Unpublished doctoral thesis, University of Utah, 1958.

BROWN J. S., CLARKE, F. R., & STEIN, L. A new technique for studying spatial generalization with voluntary responses. *J. exp. Psychol.*, 1958, **55**, 359–362.

BRUNER, J. S., & Goodman, C. D. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.

CARTER, D. E. Some effects of inter-sensory stimulation and drive on psychophysical responses. Unpublished report. Ann Arbor: Research Center for Group Dynamics, 1962.

CARTER, D. E., & ZAJONC, R. B. Studies on drive and incentive in perception: VI. Some effects of drive produced by electric stimulation on generalized responses to loudness. Technical Report No. 16, AF-49(638)367, G-4951. Ann Arbor: Research Center for Group Dynamics, 1962.

CHAMPION, R. A. Stimulus-intensity effects in response evocation. *Psychol. Rev.*, 1962, **69**, 428–449.

DORFMAN, D. D. Some effect of drive on the effective intensity of a stimulus. *Psychol. Rep.*, 1961, **9**, 87–98. (a)

DORFMAN, D. D. Some effects of drive on the perceived intensity of a stimulus. Unpublished doctoral thesis, University of Michigan, 1961. (b)

DORFMAN, D. D., PLATZ, A., & ZAJONC, R. B. Studies on drive and incentive in perception: VII. Some effects of extraneous sound and manifest anxiety on discrimination behavior. Technical Report No. 17, AF-49(638)367. Ann Arbor: Research Center for Group Dynamics, 1962.

DUFY, E. *Activation and behavior.* New York: Wiley, 1962.

ESTES, W. K. Stimulus-response theory of drive. In M. R. JONES (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: Univer. Nebraska Press, 1958. Pp. 35–68.

FUSTER, J. M. Effects of stimulation of brain stem on tachistoscopic perception. *Science*, 1958, **127**, 150.

GILBERT, G. M. Inter-sensory facilitation and inhibition. *J. gen. Psychol.*, 1941, **24**, 381–407.

HEBB, D. O. Drives and the C. N. S. (conceptual nervous sytem). *Psychol. Rev.*, 1955, **62**, 243–253.

HILGARD, E. *Theories of learning.* New York: Appleton-Century-Crofts, 1956.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

KARP, S. E. Studies on drive and incentive in perception: III. Some effects of drive on stimulus generalization obtained with a revised training procedure. Technical Report No. 8, AF-49(638)367, G-4951. Ann Arbor: Research Center for Group Dynamics, 1961.

LINDSLEY, D. C. Psychophysiology and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 44-105.

LOGAN, F. A. A note on stimulus intensity dynamism (*V*). *Psychol. Rev.*, 1954, **61**, 77-80.

LONDON, I. D. Research on sensory interaction in the Soviet Union. *Psychol. Bull.*, 1954, **51**, 531-568.

MATLIN, A. H. Studies on drive and incentive in perception: II. The effect of drive produced by proprioceptive stimulation on generalized responses to loudness and pitch. Technical Report No. 6, AF-49(638)367, G-4951. Ann Arbor: Research Center for Group Dynamics, 1960.

NEWMAN, J. R. Stimulus generalization of an instrumental response under high and low levels of drive. *Amer. Psychologist*, 1955, **10**, 459-461.

NYGAARD, J. E. Cue and contextual stimulus intensity in discrimination learning. *J. exp. Psychol.*, 1958, **55**, 195-199.

PERKINS, C. C., JR. The relation between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, **46**, 225-231.

PLATZ, A. Some effects of conditioned fear on stimulus generalization. Unpublished doctoral thesis, University of Michigan, 1962.

PLATZ, A., & ZAJONC, R. B. Effects of change in drive on stimulus generalization. Paper read at Midwestern Psychological Association, Chicago, May 1961.

RYAN, T. A. Interrelations of the sensory systems in perception. *Psychol. Bull.*, 1940, **37**, 659-698.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. Behavior theory and selective learning. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1948.

STELLER, E. The physiology of motivation. *Psychol. Rev.*, 1954, **61**, 5-22.

YAMAGUCHI, H. E. Gradients of drive stimulus (S_D) intensity generalization. *J. exp. Psychol.*, 1952, **43**, 298-304.

ZAJONC, R. B. & CROSS, D. V., Stimulus generalization as a function of drive shift. *J. exp. Psychol.*, 1964 (in press).

(Received June 7, 1963)

Psychological Review
1964, Vol. 71, No. 4, 291–313

A STRUCTURAL BALANCE MODEL OF COMMUNICATION EFFECTS[1]

N. T. FEATHER

University of New England, New South Wales, Australia

A communication model is presented in which the source, the communication, the issue, and the receiver are conceived as elements of a communication structure and attitudinal and unit relations between these elements are investigated. This basic communication structure is represented in terms of the mathematical theory of linear graphs, and Heider's principle of structural balance is invoked to specify balanced and unbalanced structures. The model is applied to interpretation of results of a sample of studies from different areas and is shown to have considerable power and generality in interpreting communication effects. Future avenues of developing the model are noted involving conceptualization of: (a) strength of relations, (b) importance of issue, (c) specific effects of imbalance, and (d) individual differences in intolerance of imbalance.

There is now a considerable body of evidence in the psychological literature to support the conclusion that a person tends to *evaluate* communications in a manner consistent with his relevant attitudes. If a communication about some issue is consistent with a person's attitude towards that issue it is more likely to be accepted than rejected. The program of research in attitude and communication by Hovland and his colleagues (Hovland, Janis, & Kelley, 1953; Hovland & Rosenberg, 1960; Sherif & Hovland, 1961) and the investigations generated by Festinger's (1957) theory of cognitive dissonance, provide evidence which is in general agreement with this conclusion. Recent studies concerned with the evaluation of evidence about the cigarette smoking-lung cancer relationship (Feather, 1963) and with judgments of the soundness or unsoundness of religious arguments (Feather, 1964) are also consistent with this conclusion.

Theoretical models have been presented which permit interpretation of this tendency for consistency between evaluation and attitude.[2] Such a tendency could, for example, be interpreted as one way of reducing cognitive dissonance (Brehm & Cohen, 1962; Festinger, 1957) or in terms of latitudes of acceptance or rejection (Sherif & Hovland, 1961). The present paper explores another possible approach which is based on Heider's (1946, 1958) theory of balanced structures and which, in particular, draws upon the reformulation of this theory by Cartwright and Harary (1956), using concepts from the mathematical theory of linear graphs. This approach therefore considers the relationship of evaluation to attitude in terms of a structural balance model. Such a model also has implications about mutual attitudes between the source and the receiver of the communication which may be reflected in judgments about *credibility*. It also provides an analysis of the conditions under which one might expect *attitude change* to occur. It may therefore be viewed as a relatively general model of communication effects. We will now turn to a statement of this model and then to an examination of some experimental evidence pertaining to it.

---

[1] Sections of this paper were written when the author was Visiting Scientist and Research Associate at the University of Michigan under the sponsorship of the Institute of Science and Technology and National Science Foundation Project GS-9.

[2] The main alternative models are discussed in recent papers by Osgood (1960), Zajonc (1960), and Brown (1962).

COMMUNICATION STRUCTURE

The communication structure with which we are concerned consists of the following *elements:* a source (S), a receiver (R), a communication (C), and an issue (I). The source is delivering a communication about an issue to a receiver. The communication may be verbally presented or written, but it takes some position on the issue. This is the basic communication structure to be considered and is an abstraction from such diverse situations as a lawyer presenting an argument about the innocence of his client to a juryman, a minister preaching about religion to a member of his congregation, a newspaper editor writing about racial integration for a reader, a therapist giving his interpretation of a particular symptom to a client, a politician outlining a proposal of agricultural policy to a member of his electorate, and a teacher presenting a lesson on the American Civil War to a pupil.

It is assumed that we know certain *attitudinal relations* between elements of the communication structure. Specifically, we assume that the attitudes of the source, the communication, and the receiver towards the issue are known. These attitudinal relations are either positive or negative. At present no allowance is made in the model for variation in the *strength* of the attitude. Only the *sign* of the attitude is considered. The various situations presented above may be expanded to take account of the possible attitudinal relations. For example, the newspaper editor may be in favor of racial integration, his editorial may advocate racial integration, and his reader may be opposed to racial integration. The politician may be personally opposed to a change in agricultural policy, his speech may, in fact, advocate a change in agricultural policy, and his audience may favor a change in agricultural policy.

The elements and relations in the above communication structure are coordinated to concepts from the theory of linear graphs (Cartwright & Harary, 1956). In particular, the elements are coordinated to *points* of the graph, and the relations to *directed lines of the graph*. The sign of the relation is represented by the *sign of the directed line*. The resulting graph is called a *signed digraph* or *s-digraph*. For the communication structure under discussion, there are eight distinct s-digraphs which include all possible combinations of positive and negative attitudinal relations. These eight s-digraphs are presented in Figure 1 employing the convention that solid lines are positive and dashed lines negative. The arrows on the lines of each s-digraph indicate direction and represent the asymmetry of the attitudinal relations in that source, communication, and receiver are all assumed to have attitudes towards the issue, but the issue cannot have attitudes towards them.

In Figure 1 we represent at one extreme, in s-digraph a, the structure in which source, communication, and receiver are all in favor of the issue The positive line $\overrightarrow{SI}$ means that the source of the communication has a positive attitude or is favorably disposed towards the issue. Correspondingly, the positive lines $\overrightarrow{RI}$ and $\overrightarrow{CI}$

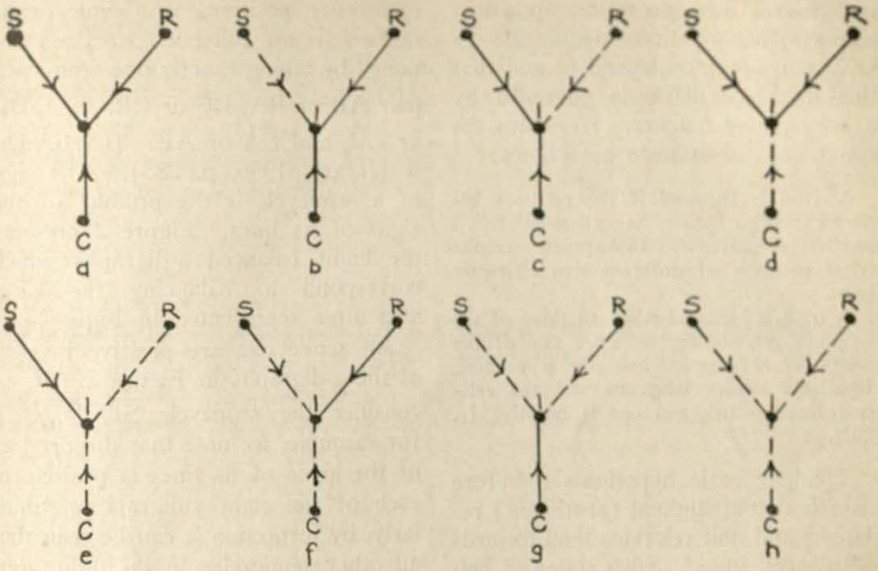


FIG. 1. Signed digraphs representing the eight possible combinations of positive and negative attitudinal relations of source (S), receiver (R), and communication (C) to issue (I).

mean that both receiver and communication are favorably disposed towards the issue. The structure in s-digraph a could, for example, represent the situation in which a minister with a positive attitude towards religion delivers a sermon favoring religion to a member of his congregation who favors religion. At the other extreme, we represent in s-digraph h the structure in which source, communication, and receiver are all opposed to the issue. The negative lines $\overrightarrow{SI}$, $\overrightarrow{RI}$, and $\overrightarrow{CI}$ represent these uniformly negative attitudinal relations. The structure in s-digraph h could, for example, represent the situation in which a conservative politician opposed to socialized medicine argues against socialized medicine to an audience also opposed to this policy. Between the extreme structures of s-digraph a and s-digraph h are intermediate structures involving both positive and negative attitudinal relations.

STRUCTURAL BALANCE

Each of the eight structures in Figure 1 represents, in the form of an s-digraph, three attitudinal relations between elements of the structure, viz., source to issue, communication to issue, and receiver to issue. The attitudinal relations of source to communication, of receiver to communication, and of source to receiver (and vice versa) are unknown. It is assumed that these relations may be predicted by applying Heider's (1958) principle of balance.

Heider proposes this principle in an analysis of structures involving two or more of the following entities:— P (one person), O (another person), and X (an impersonal entity). The entities may be related *attitudinally* (e.g., P likes X, O does not like P), or in terms of a *unit relationship* (e.g., P owns X, P and O live together). Heider (1958), reflecting the general field-theoretical approach, assumes that relations within the P-O-X unit are interdependent and that this interdependence is governed by a *principle of balance*. He states the conditions for balance as follows:

> A dyad is balanced if the relations between the two entities are all positive . . . or all negative. . . . Disharmony results when relations of different sign character exist.
>
> A triad is balanced when all three of the relations are positive or when two of the relations are negative and one is positive. Imbalance occurs when two of the relations are positive and one is negative [p. 202].

Heider's basic hypothesis therefore asserts that attitudinal (sentiment) relations and unit relations tend towards a balanced state.[3] Such states of balance tend to be preferred over disharmony. Existing relations may induce new relations in the direction of a balanced state, and when a structure is unbalanced pressures will arise towards balance. Transformation from a state of imbalance to a state of balance may occur by means of changes either in the attitudinal relations or the unit relations or both. If such changes are not possible the state of imbalance will produce tension.

Cartwright and Harary (1956) reformulate the concept of a balanced structure using the mathematical theory of linear graphs and, in so doing, are able to resolve some ambiguities in Heider's (1946) presentation and to generalize the principle to more complex structures. In particular, they assume that an s-digraph is completely balanced if all its *semicycles* are positive. A semicycle is defined as "a collection of lines obtained by taking exactly one from each pair $\overrightarrow{AB}$ or $\overrightarrow{BA}$, $\overrightarrow{BC}$ or $\overrightarrow{CB}$, . . . , $\overrightarrow{DE}$ or $\overrightarrow{ED}$, and $\overrightarrow{EA}$ or $\overrightarrow{AE}$. [Cartwright & Harary, 1956, p. 283]." The sign of a semicycle is the product of the signs of its lines. Figure 2 presents the eight balanced s-digraphs which correspond to balancing the eight structures represented in Figure 1.

All semicycles are positive in each of the s-digraphs in Figure 2. If we consider the semicycle $\overrightarrow{SI}$, $\overleftarrow{IC}$, $\overleftarrow{CS}$, for example, we note that the product of the signs of its lines is positive in each of the eight s-digraphs. Similarly, by inspection it can be seen that all other semicycles in the s-digraphs are positive. Therefore all of the s-digraphs in Figure 2 are completely balanced.

In each s-digraph the S-I-R unit involves three semicycles: $\overrightarrow{SR}$, $\overrightarrow{RS}$; $\overrightarrow{SR}$, $\overrightarrow{RI}$, $\overleftarrow{IS}$; and $\overleftarrow{SR}$, $\overrightarrow{RI}$, $\overleftarrow{IS}$. This unit corresponds to the P-O-X unit of Heider for the special case where one person is the source of a communication, the other person is the receiver of the communication, and the impersonal entity is the issue which the communication concerns. It will be noted that symmetric relations are permitted between points S and R, i.e., the source and the receiver may have attitudes towards one another. The relations between points S and C, and R and C are, however, asymmetric since it is assumed that while both source and receiver can have attitudes towards the communication, the communication does not have attitudes towards them.

The structures represented in Figure 2 permit predictions about attitudinal relations of source to communication, receiver to communication,

---

[3] Cartwright and Harary (1956, p. 278) provide the following examples of units balanced according to the above conditions: P likes something he made; P likes what his friend likes; P dislikes what his friend dislikes; P likes what his enemy dislikes; and P's son likes what P likes.

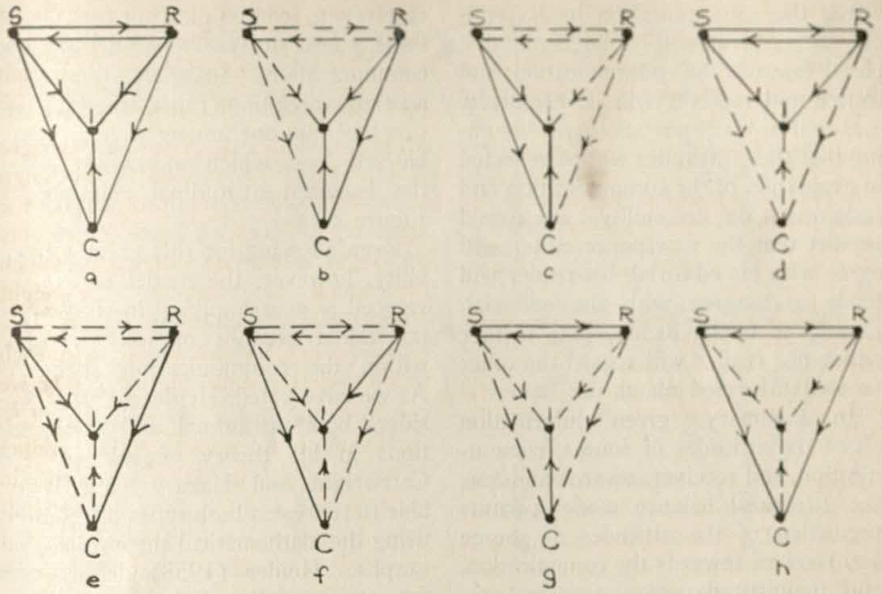


FIG. 2. Signed digraphs in which the eight s-digraphs of Figure 1 are completely balanced for attitudinal relations according to Heider's principle.

and source to receiver (and vice versa), given the information contained in the s-digraphs of Figure 1 and the assumption of structural balance. Thus, knowing the attitudes of source, communication, and receiver towards the issue with which the communication is concerned, it is possible, by employing the balanced s-digraphs in Figure 2, to predict *four* attitudinal relations: the attitude of source to communication, the attitude of receiver to communication, the attitude of source to receiver, and the attitude of receiver to source. These predictions concern the sign of the attitude (whether it is positive or negative) but not its strength. We assume that the attitude of source or receiver to the communication is reflected in the way he *evaluates* the communication, i.e., whether he accepts or rejects it, agrees or disagrees with it, is convinced or not convinced by it. Further, we assume that the attitude of source to receiver and vice versa may be reflected in the evaluation made of his *credibility,* i.e., whether he is perceived as well informed, trustworthy, etc.

For example, let us consider the case of the newspaper editor in favor of racial integration whose editorial advocating racial integration is read by a person opposed to this policy. We assume also that the editor knows the reader's attitude to racial integration and vice versa. This communication structure is represented in Figure 1 by s-digraph c, and predictions based on the structural balance model may be obtained by referring to the corresponding balanced s-digraph c in Figure 2. There it is apparent that the directed line $\overrightarrow{SC}$ is positive, and that the directed lines $\overrightarrow{RC}$, $\overrightarrow{SR}$, $\overrightarrow{RS}$ are all negative. We would therefore predict that, for this particular structure, the source will be positively disposed towards the communication he is sending, the receiver will be negatively disposed towards the communication, and source and receiver will be negatively disposed towards one another. Assuming that these attitudes will be reflected in evaluation of the communication and judgments of credibility, we would predict that the newspaper editor will agree with his editorial, his reader will tend to disagree with the editorial, finding it unconvincing, and neither editor nor reader will regard the other as well informed about the issue.

In summary, given information about the attitudes of source, communication, and receiver towards an issue, the structural balance model permits prediction of the attitudes of source and receiver towards the comunication, and the attitudes of source and receiver towards one another. These attitudes are assumed to be reflected in evaluation of the communication and in judgments of the credibility of the source or the receiver.

EXTENDED MODEL

To this stage we have limited our statement of the model to the particular case where attitudinal relations of source, communication, and receiver towards the issue are specified and the attitudinal relations of source and receiver to communication, and source and receiver towards each other, are predicted. But it should be apparent that the type of approach described above can be extended to deal with cases where other sets of attitudinal relations between elements of the communication structure are specified and the remaining attitudinal relations are predicted using the principle of structural balance. For example, given the attitudinal relations of receiver to source, receiver to issue, and communication to issue we can use Figure 2 to predict the attitudinal relations of receiver to communication, source to issue, source to receiver and source to communication. In short, the set of attitudinal relations represented in Figure 1 is only one among several possible sets from which one could predict the balanced attitudinal relations in Figure 2.

Even allowing for this greater flexibility, however, the model so far developed is oversimplified in that it has not taken account of *unit relations* within the communication structure. As we have noted, Heider (1958) considers both attitudinal and unit relations in his theory of balance, and Cartwright and Harary (1956) are able to represent both types of relations using the mathematical theory of linear graphs. Heider (1958) defines unit formation as follows: "separate entities comprise a unit when they are perceived as belonging together. For example, members of a family are seen as a unit; a person and his deed belong together [p. 176]." Cognitive unit formation would refer to such specific relations between entities as similarity, proximity, ownership, causality, or belonging.

In the present context, we will consider unit relations between source and communication, between receiver and communication, and between source and receiver.[4] Figure 3 presents the eight balanced s-digraphs in which both attitudinal and unit relations are

---

[4] Unit relations of source to issue, receiver to issue, and communication to issue could also be included in Figure 3 but have not been so included since we are not sure just what these unit relations might represent. Positive unit relations of source and receiver to the issue might be represented in some sense of "ownership" of the issue, or appropriateness of the issue to the person. A positive unit relation of communication to issue might be reflected in a judgment that the communication is pertinent or relevant to the issue.

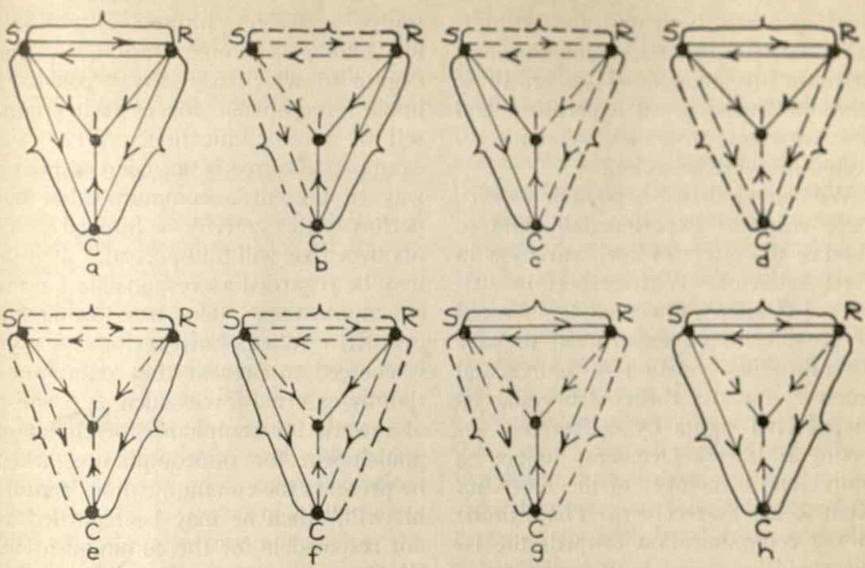


FIG. 3. Signed digraphs representing the eight communication structures in which all semicycles involving attitudinal and unit relations are completely balanced.

represented. A positive unit relation is represented by a solid brace; a negative unit relation is represented by a dashed brace. Following Heider we assume that a positive unit relation implies unit formation and that a negative unit relation implies a resistance to unit formation.

The s-digraphs in Figure 3 are of Type 2 (Cartwright & Harary, 1956), since *two* different relations (unit and attitudinal) are defined on the same set of elements. In contrast, the s-digraphs in Figures 1 and 2 are of Type 1. Following Cartwright and Harary (1956, p. 285) we define an s-digraph of Type 2 as balanced if all its semicycles are positive.[5] All semicycles are positive in each of the s-digraphs in Figure 3. For example, if we consider the semicycle consisting of $\overrightarrow{RI}$, $\overleftarrow{IC}$, and the unit relation CR in each of the eight s-digraphs, we note that in all cases the product of the signs of its lines is positive. Similarly, by inspection it can be seen that all other semicycles in the s-digraphs are positive, irrespective of whether they involve only attitudinal relations, unit relations, or both types of relation. Therefore, all of the s-digraphs in Figure 3 are completely balanced.

The model presented in Figure 3 has considerable predictive power. In fact, it is possible to predict *all* the relations represented in each s-digraph in Figure 3 providing *three* relations are known which involve the *four* elements (S, C, I, R) of the communication structure. For example, knowing the attitudinal relation of S to I, the attitudinal relation of R to I, and the unit relation of S to C, one can predict all other relations within the communication structure. It is obvious, however, that an adequate test of the above model will depend upon the possibility of coordinating the theoretical concepts to empirical operations. In particular, it is necessary to have some operational specification of both unit and attitudinal relations for the communication situation investigated. How are these relations varied in an experiment and how are changes in these relations reflected in overt behavior?

---

[5] Other assumptions are possible (e.g., Morrissette, 1958).

We believe that it is possible to vary these relations experimentally and to observe the effect of such variation in overt behavior. With respect to attitudinal relations, for example, it would be possible in an experiment to vary the attitudinal relation of source and receiver towards the controversial issue of civil rights by deliberately selecting a liberal Democrat to act as source and a member of the Ku Klux Klan to act as receiver. The attitude of the communication towards the issue could be varied by using a strong pro-civil rights message in one situation and an anti-civil rights message in another. If we were predicting the attitudinal relations of source and receiver towards the communication and the attitudinal relations of source and receiver towards one another, we could observe the way each evaluates the communication (e.g., acceptance versus rejection) and the evaluative reactions (positive or negative) of one toward the other.

Similarly, one could operationally specify a positive unit relation between source and receiver by selecting subjects who are, for example, members of a strong cohesive group. A negative unit relation could be specified by choosing as source and receiver subjects who are members of rival factions.

The coordination of unit relations between source and communication and between receiver and communication to experimental operations is less obvious but one possibility will be discussed in some detail because of its importance in the analysis of communication effects. Within an experiment it is possible to vary the degree to which a source considers himself *responsible* for sending a communication, and the degree to which a receiver considers himself responsible for exposing himself to a communication. When, for example, a source is not coerced in any way to present a communication but performs the activity voluntarily, "of his own free will and accord," then he may be regarded as responsible for the communication. He has ownership over it. When, however, the source is coerced to present the communication by external forces such as promise of reward for compliance or threat of punishment for noncompliance, when he presents the communication "against his will," then he may be regarded as not responsible for the communication. We assume that responsibilty of a source for a communication may be represented by a *positive* unit relation. A *negative* unit relation would imply that the source is not responsible for the communication and, in fact, disowns it.

The concept of responsibility has some antecedents in the psychological literature. In the particular context of achievement behavior, for example, a related variable is the degree to which a person believes that he personally *controls* his success or failure, the degree to which he considers that success or failure is due to his own skill and efforts and not to the operation of chance or external agencies (cf. Feather, 1959). Similarly, in the communication situation the source's responsibility would depend upon the degree to which he is able to control both his acceptance or rejection of the role of source and his continuance in that role. Where external forces are such that they primarily determine his role as source then he may not feel responsible for the communication. Similar ideas are presented in a recent publication by Brehm and Cohen

(1962) where they suggest that an important factor that may control the magnitude of dissonance in a situation is the degree of *volition* involved. They define degree of volition as:

> the extent to which a person feels that he controls his own behavior. . . . Volition implies not only initiation and selection of behavior but also responsibility for its consequences [p. 201].

Similarly, one can represent a receiver's responsibility for exposing himself to a communication by a unit relation. A positive unit relation would represent the case where the receiver has actively and voluntarily exposed himself to a communication. For example, a person of his own accord attends a lecture advocating the virtues of Communism. In contrast, a negative unit relation would represent the case where the receiver is coerced by external forces to receive the communication. For example, a captured soldier is forced by threat of punishment to listen to a lecture extolling the virtues of Communism. Many communication studies in psychology involve some degree of coercion since the receiver of a communication is often paid to participate as a subject. Volunteer subjects, knowing the conditions of the experiment and not induced to participate for monetary reward, would presumably feel more responsible for their commitment to the receiver's role.

The examples given above of possible ways of operationally specifying attitudinal and unit relations within a communication situation are not meant to be exhaustive. These relations can probably be specified in a variety of ways and it is an empirical problem to determine which operations best fit the model. For example, in order to induce a positive attitudinal relation between receiver and source, one might introduce the source as a very warm and likable person, or as a trustworthy and well-informed authority. In line with the above discussion, one might represent by a positive unit relation the situation where a receiver has voluntarily exposed himself to a particular source. One might consider "assimilation" and "contrast" effects in a receiver's reactions to a communication as reflecting positive and negative unit relations, respectively. Many of these operations have already been investigated in studies of communication and attitude change.[6]

We will now consider a selection of these experimental studies, representing different theoretical approaches, to show how they may be interpreted in terms of the structural balance model developed above.

SOME EXPERIMENTAL STUDIES

The studies to be discussed are classified as follows: (*a*) Studies in which no deliberate attempt is made to induce imbalance in the communication structure; (*b*) Studies in which there is a deliberate attempt to induce imbalance in the communication structure. Both sets of studies are relevant to the validity of the structural balance model.

Studies without Induced Imbalance

In his section we will consider three investigations, one of our own studies concerned with the evaluation of a communication, a study by Kelman and Hovland (1953) concerned with the effects of variations in source credibility, and a study by Hovland, Harvey, and Sherif (1957) concerned with assimilation and contrast effects in reactions to communication.

Feather. This study was designed to investigate attitudinal relations involving source, receiver, and communication, given the attitudinal relations represented in Figure 1. Subjects were involved in a role playing situation in which they were required to identify with either the source or the receiver of a communication about a controversial issue and to predict his reactions. Each subject was given information about the attitudes of the source, the receiver, and the communication towards the issue. This information was presented in the eight possible combinations of attitudinal relations represented in the s-digraphs of Figure 1, using a random order of presentation. All subjects were required to predict whether the source and receiver would tend to agree or disagree with the communication. Subjects identifying with the source were required to predict whether the source would tend to regard the receiver as well informed or not well informed about the issue. Subjects identifying with the receiver were required to predict whether the receiver would tend to regard the source as well informed or not well informed about the issue. Thus, the attitudinal relations of source and receiver towards the communication were assumed to be reflected in their *evaluation* of the communication. The attitudinal relations of source and receiver towards one another were assumed to be reflected in judgments of *credibility* of source or receiver.

---

[6] Newcomb's (1959, 1961) work using the A-B-X model is of particular interest in indicating how balanced structures may develop over a period of time.

The results of this experiment provided strong support for the structural balance model. They indicated that a person's evaluation of a communication will tend to be *positive* when his attitude towards the issue is of the same sign as the communication's attitude towards the issue, and *negative* when these attitudes are of opposite sign. Similarly, the results showed that a source (or receiver) will tend to judge a receiver (or source) as well informed about an issue when he knows that they both have attitudes towards the issue which are of the same sign, and not well informed when he knows that they both have attitudes towards the issue which are of opposite sign. The results also indicated that a person's judgments of credibility are likely to be relatively less balanced than are his evaluations of the communication. We believe that this relative difference reflects the fact that the attitudinal relation of source to receiver (or vice versa) may be manifested in a variety of ways depending on the situation. For example, a receiver may have the same attitude as the source towards a certain issue but he may have reason to doubt that the source's attitude is based on reliable evidence. In this case, the predicted positive attitudinal relation of receiver to source may be manifested in some form other than a judgment that the source is credible (e.g., in a certain liking or sympathy for the source, a positive attraction for the source despite the realization that he may not be well informed). Such alternative manifestations of an attitudinal relation are probably less likely where the relation of person to communication is concerned. Here the relation is more likely to be expressed in terms of agreement or disagreement with the communication.

Kelman and Hovland (1953). We will limit our discussion of this study to that aspect concerned with the effects of variation in source credibility on evaluations of a communication. High school students listened to a recording of an educational program in the course of which a guest speaker gave a talk favoring lenient treatment of juvenile delinquents. To vary the credibility of the source the speaker was introduced in three different ways. In the *positive credibility* condition he was identified as the presiding judge in a juvenile court, author of several books on delinquency, etc. In the *neutral credibility* condition he was introduced as a member of the studio audience, chosen at random. In the *negative credibility* condition he was presented as a person with a shady past, a near-delinquent himself, out on bail on a dope-peddling charge, etc. Among questions asked of subjects after the speech was the following: "Was the guest's presentation fair or was it one-sided?" The results indicated that the percentage of subjects who thought the communicator's presentation was "completely fair" or "fair" decreased substantially as the credibility of the communicator decreased.

The results may be understood in terms of the two balanced s-digraphs in Figure 4. In s-digraph a the positive attitudinal relation of receiver to source represents the positive credibility of the source, the positive unit relation of source to communication assumes that the source is seen as responsible for the communication, and the positive attitudinal relation of receiver to communication balances the S-C-R triad. The latter attitudinal relation is assumed to be reflected in the subject's judgment that the communication has been fairly presented. In a corresponding way, the negative attitudinal relation of receiver to communication in s-digraph b balances the S-C-R triad, which represents the case in which a negatively credible source is regarded as responsible for a one-sided communication.

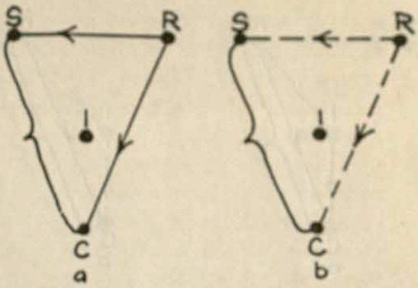


FIG. 4. Balanced s-digraphs representing the effect of variation in source credibility on receiver's evaluation of communication.

Hovland, Harvey, and Sherif (1957). In this investigation each subject was first required to indicate his stand on a controversial issue (prohibition and repeal in a "dry" state) by checking which of nine representative statements about the issue best corresponded to his own position. These statements about the issue ranged from an extreme "dry" stand to an extreme "wet" stand. Each subject also checked those statements which he did not find objectionable from his point of view, those statements which he did find objectionable, and those statements which were neither acceptable nor unacceptable to him. This procedure provided information about his *latitude of acceptance* (the range of statements tolerated) and his *latitude of rejection* (the range of statements rejected). Subsequently a wet (repeal) communication was presented to extreme dry subjects and unselected subjects, a dry (prohibition) communication was presented to extreme wet and unselected subjects, and a moderate communication was presented to wet, dry, and unselected subjects. Following the communication the nine statements were again presented to each subject to secure his stand, his latitude of acceptance, and his latitude of rejection concerning the issue. His reactions to the communication were also obtained through ratings on like-dislike, reasonable-unreasonable, biased-unbiased, and propaganda-fact dimensions. Subjects who received the moderate communication also checked on a rating scale ranging from extreme dry to extreme wet positions what they thought to be

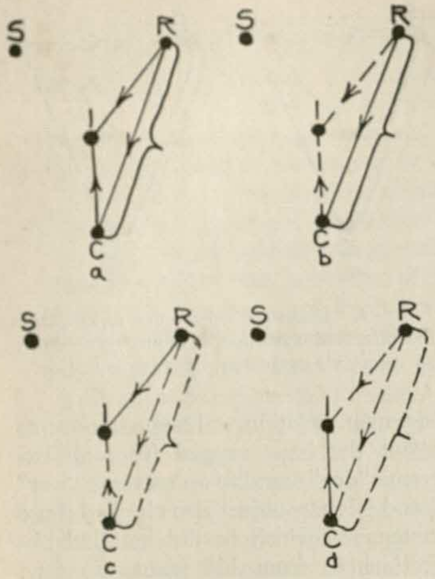


FIG. 5. Balanced s-digraphs representing cases where communication is within receiver's latitude of acceptance (s-digraphs a and b) and within his latitude of rejection (s-digraphs c and d).

the stand taken in the communication itself.

The results indicated that there was an extremely close relationship between the person's own stand concerning the issue and his evaluation of the communication. Hovland, Harvey, and Sherif (1957) conclude that,

> The data provide quantitative information . . . that individuals who are in favor of the opinion advocated will consider the communication fair and unbiased, but that those with an opposed stand will regard an identical communication as propagandistic and unfair [p. 247].

In discussion of this result they suggest that when a communication falls within the latitude of acceptance it will tend to be judged as fair and unbiased. As it is perceived as moving more and more into the latitude of rejection there is an increasing tendency for it to be considered unfair and propagandistic. The results of this investigation also showed that subjects with extreme stands on the issue tended to perceive the moderate communication as further removed from their own stand than it actually was (a "contrast" effect). Subjects with stands close to the position advocated in the moderate communication tended to perceive the communication as closer to their own stand than was actually the case (an "assimilation" effect).

The results of this study may be interpreted in terms of the balanced s-digraphs in Figure 5. Since no information is provided in the study about the credibility of the source, no relation is introduced between S and R. We assume that a communication falls within a person's latitude of acceptance when the attitudinal relations of both receiver and communication to the issue are of the same sign, and within a person's latitude of rejection when these relations are of opposite sign. Considering attitudinal relations, Figure 5 shows that the R-I-C triad would be balanced by a positive attitudinal relation when the communication falls within the individual's latitude of acceptance (s-digraphs a and b) and by a negative attitudinal relation when the communication falls within the individual's latitude of rejection (s-digraphs c and d). The results are consistent with these predicted attitudinal relations, assuming that a positive attitudinal relation of receiver to communication may be reflected in a judgment that the communication is fair and unbiased, and that a corresponding negative attitudinal relation may be reflected in a judgment that the communication is unfair and propagandistic. The "assimilation" and "contrast" effects obtained in this study indicate that the receiver may misinterpret the attitudinal relation of the communication to the issue when the communication is somewhat moderate or intermediate in its stand. With respect to the present analysis the assimilation effect may be interpreted as a tendency for the receiver to misperceive the attitudinal relation of communication to issue so that it is seen as closer to his own attitude about the issue than is actually the case, when the communication falls within his latitude of acceptance (s-digraphs a and b). The contrast effect may be interpreted as a tendency for the receiver to misperceive the attitudinal relation of communication to issue so that it is seen as more distant from his own attitude than is actually the case, when the communication falls within his latitude of rejection (s-digraphs c and d). It may also be possible to consider the assimilation effect as a manifestation of a positive unit relation between receiver and communication (s-digraphs a and b) and the contrast effect as a manifestation of a negative unit relation between receiver and communication (s-digraphs c and d).

It will be noted that the above interpretation does not enable precise representation of the size of the *discrepancy* between the stand taken by the receiver of a communication and the position taken in the communication. Discrepancy is considered as a difference in the *sign* of the attitudinal relations. A more refined analysis of this type of experiment would involve taking account of the differing *strengths* of the attitudinal relations as well as the differing signs. As we will later indicate, this is one important direction in which the structural balance model needs to be developed.

Studies with Induced Imbalance

In this section we will first consider a discussion by Hovland, Janis, and Kelley (1953) concerned with the effects of varying source credibility. This will be followed by a discussion of Tannenbaum's investigation of attitude change (Osgood & Tannenbaum, 1955) and Festinger and Carlsmith's (1959) study of the effects of presenting an attitude-discrepant communication.

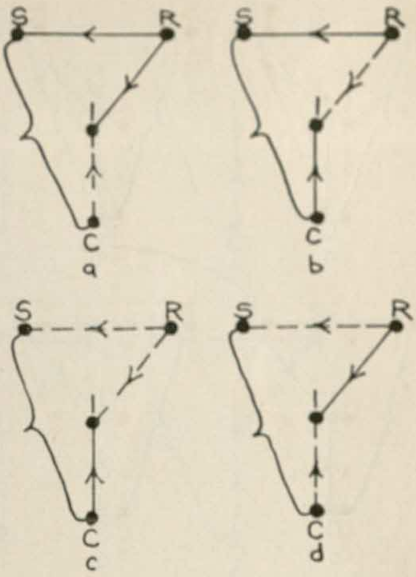


FIG. 6. Unbalanced s-digraphs (a and b) and balanced s-digraphs (c and d) representing the case where attitudinal relations of receiver and communication to issue are at variance.

Hovland, Janis, and Kelley (1953). These authors are particularly concerned (pp. 40–46) with the situation in which a positive communicator presents a communication which, on its own merits, would tend to be rejected. The two possible forms of this structure are represented in s-digraphs a and b in Figure 6.

These two s-digraphs represent situations where the source is responsible for a communication (positive unit relation), where the receiver is positively disposed towards the source (positive attitudinal relation), but

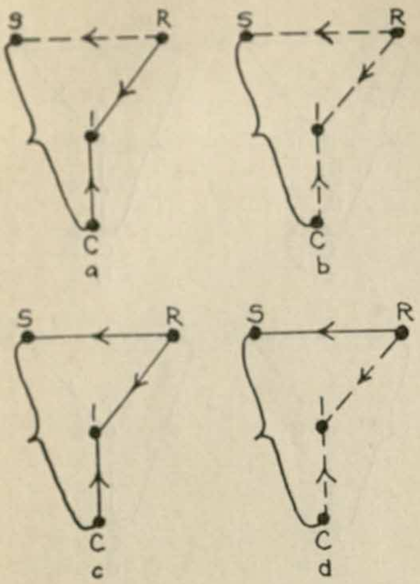


FIG. 7. Unbalanced s-digraphs (a and b) and balanced s-digraphs (c and d) representing the case where attitudinal relations of receiver and communication to issue are not at variance.

where the attitudes of receiver and communication towards the issue are at variance. These structures are both in a state of imbalance since not all the semicycles can be positive at the one time. For example, the structure corresponding to semicycle $\overrightarrow{RC}$, $\overrightarrow{CI}$, $\overleftarrow{IR}$ in s-digraph a would be balanced if $\overrightarrow{RC}$ were negative, since the semicycle would then be positive in sign. However, a negative directed line $\overrightarrow{RC}$ would mean that the semicycle involving $\overleftarrow{SR}$, $\overrightarrow{RC}$, and the positive unit relation SC would be negative, i.e., the structure corresponding to this semicycle would be out of balance. The discussion by Hovland, Janis, and Kelley (1953) can be seen as indicating how such imbalances are resolved. The possibilities they consider, which include dissociating the source and the communication, reinterpreting the meaning or attitude of the communication, and a change in the receiver's attitude towards the source, can readily be interpreted using the structural balance principle. For example, the receiver might dissociate the source from the communication. This dissociation would be more likely to occur the more *ambiguous* the evidence that the source is responsible for the communication. Dissociating the source from the communication may be represented by a negative unit relation between source and communication. It is evident that, if the unit relations in s-digraphs a and b were negative rather than positive, these s-digraphs would then be in balance. This analysis also suggests that a receiver probably expects that the attitudes of both source and communication towards an issue should tend to be in harmony when the source is obviously responsible for the communication. Where, however, the receiver sees the attitudes of source and communication towards the issue as at variance, he may perceive the source as not really responsible for the communication.

Figure 6 also presents s-digraphs c and d which represent communication structures in which complete balance is possible. In these two situations the attitudes of the receiver and the communication towards the issue are again at variance, but the receiver is negatively disposed towards the source. There are four other structures, corresponding to those represented in Figure 6, in which the attitudes of receiver and communication towards the issue are *not* at variance. These structures are balanced providing the receiver perceives the source as credible, i.e., there is a positive attitudinal relation of receiver to source. In Figure 7, s-digraphs a and b are unbalanced; s-digraphs c and d are balanced.

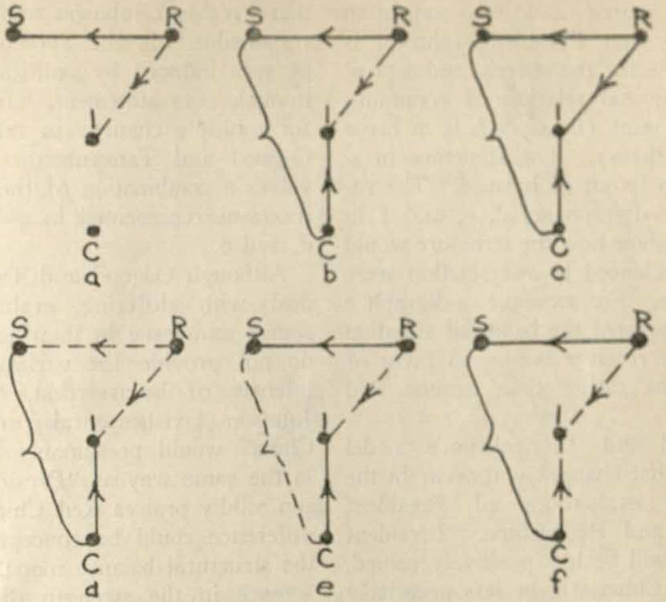


FIG. 8. Balanced s-digraphs (c, d, and e) representing predicted changes in relations in unbalanced s-digraph b according to the principle of congruity.

Osgood and Tannenbaum (1955). In his test of the congruity model Tannenbaum first required subjects to evaluate six cognitive objects consisting of three sources (Labor Leaders, *Chicago Tribune,* and Senator Robert Taft) and three objects (Legalized Gambling, Abstract Art, and Accelerated College Programs). Communications consisting of "highly realistic newspaper stories" were then constructed involving the following source-concept pairs: Labor Leaders–Legalized Gambling; *Chicago Tribune*–Abstract Art; Senator Robert Taft–Accelerated College Programs. Each group of subjects read a favorable newspaper story for one of these pairs, an unfavorable story for a second pair, and had no story at all for the third pair. After reading the stories they again evaluated the six sources and concepts. The results showed that changes in evaluations following exposure to the communications were generally consistent with predictions from the congruity model. For example, concepts linked to a highly valued source by a favorable assertion tended to be more highly valued.

The procedure used in this study may be represented by using the s-digraphs in Figure 8. Let us suppose, for example, that in an experiment a subject gives a high positive evaluation to President Johnson (the source), and a high negative evaluation to Red China (the issue). This situation is represented in s-digraph a by a positive attitudinal relation of receiver to source and a negative attitudinal relation of receiver to issue. The subject then reads a newspaper account stating that President Johnson has made a speech (the communication) praising Red China. This new information is represented in s-digraph b by a positive unit relation between source and communication signifying that President Johnson is responsible for the speech, and a positive attitudinal relation of communication to issue (the speech is in favor of Red China). The structure in s-digraph b is out of balance. The remaining s-digraphs c, d, e, and f in Figure 8 show how the structure would become balanced if *one* relation were to change. For example, s-digraph c would represent the balanced situation where the receiver is now in favor of Red China rather than against Red China.

Osgood and Tannenbaum's model predicts that changes will occur in the receiver's evaluations of President Johnson and Red China. President Johnson will be less positively valued, and Red China will be less negatively valued. These predicted changes would be based on the congruity formula. They would then be corrected to take account of the amount of "incredulity" involved in the situation, and for the fact that the assertion applies to the issue rather than to the source. With respect to the structural balance model this approach is equivalent to saying that a number of relations may change in the communication structure to produce congruity. Typically, change would not be restricted to one relation only. Changes will tend to occur in the attitudinal relations of receiver to source and receiver to issue. These predicted changes would have to take into account the "assertion constant [Osgood & Tannenbaum, 1955, p. 50]" and reactions of incredulity. Changes in the attitudinal relations of receiver to issue and receiver to source are represented in s-digraphs c and d in Figure 8. The reaction of incredulity may be represented in s-digraph e by a negative unit relation between source and communication. For example, the receiver may assume that President Johnson was not really responsible for the speech but that he was induced by political pressure to make the statement. In allowing for multiple changes in relations the Osgood and Tannenbaum model involves a combination of the balancing reactions represented in s-digraphs c, d, and e.

Although Osgood and Tannenbaum deal with differing evaluations of source and issue in their model, they do not provide for variation in the intensity of the assertion. "President Johnson lavishes praise upon Red China" would presumably be treated in the same way as "President Johnson mildly praises Red China." This difference could be conceptualized in the structural balance model as a difference in the strength of the attitudinal relation of communication to issue although, as indicated previously, the present model has yet to be developed to deal with differences in the strength of relations. It is interesting to note that s-digraph b would be balanced if the attitudinal relation of communication to issue were negative (as in s-digraph f). This type of effect is likely to occur in situations where the communication is sufficiently ambiguous to allow misinterpretation by the receiver. A much qualified statement by President Johnson which is, in fact, mildly favorable towards Red China may be interpreted as antagonistic to Red China by the receiver.

Finally, s-digraphs a and b in Figures 6 and 7 represent incongruous situations in Osgood and Tannenbaum's sense. Since, according to our assumptions these s-digraphs are unbalanced, incongruity here corresponds to structural imbalance and both models would predict that changes would occur in the direction of congruity or balance. The s-digraphs c

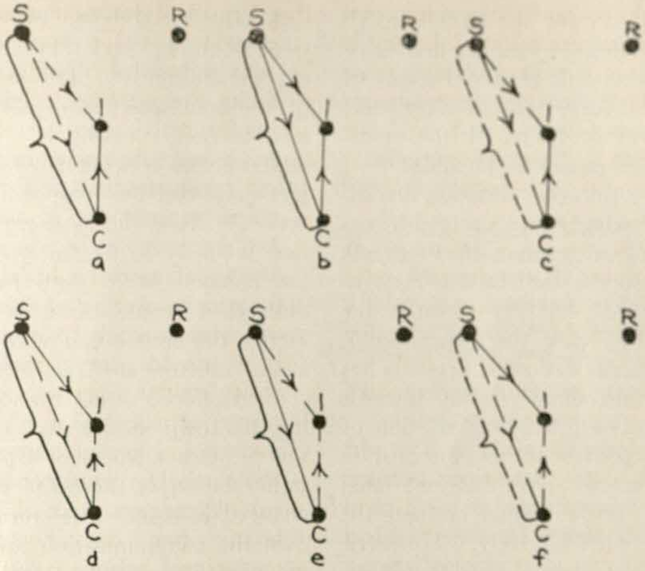


FIG. 9. Unbalanced s-digraphs (a and d) and corresponding balanced s-digraphs (b, c, and e, f) showing changes in attitudinal and unit relations.

and d in Figures 6 and 7 are balanced according to our assumptions. Even here, however, Osgood and Tannenbaum would expect changes to occur in the attitudinal relations of receiver to source and receiver to issue if these relations were initially different in strength. The structural balance model at present makes no provision for such changes. Their conceptualization would depend upon further development of the structural balance model to take account of relations differing in intensity.[7]

Festinger and Carlsmith (1959). In this experiment subjects were required to perform a boring and tedious task and then to tell the "next subject" that the task was interesting and enjoyable. Some subjects were offered $1.00 to engage in this attitude–discrepant behavior, others were offered $20.00. After they had made the discrepant statements, subjects rated how interesting and enjoyable they found the experiment. A control group of subjects, not required to make attitude–discrepant statements, also rated the experiment for how enjoyable it was. Festinger and Carlsmith found that the control group tended to rate the experiment as unpleasant. The mean rating from subjects in the $1.00 experimental condition was significantly higher than that of the control group, and definitely on the positive side of the scale. The mean rating from subjects in the $20.00 experimental condition was only slightly above the mean for the control condition, and was significantly

[7] The example discussed in this section would be conceptualized as an unbalanced +p− bond in Rosenberg and Abelson's (1960) balance model. Such a bond could also be represented by the unbalanced s-digraph b in Figure 8. Rosenberg and Abelson's model focuses upon the signs of relations while Osgood and Tannenbaum's model also considers the strength of relations. Brown (1962) notes a number of other important differences between the two models.

below the mean for the $1.00 experimental condition. Festinger and Carlsmith interpret this result in terms of dissonance theory, but an alternative interpretation in terms of the structural balance model is possible.

The communication situation investigated by Festinger and Carlsmith corresponds to s-digraph a in Figure 9, where the source initially has a negative attitudinal relation towards the issue (in this case the experiment), and the communication he presents has a positive attitudinal relation towards the issue. The attitudinal relation of source to receiver is not specified in the experiment. We assume that initially the source would evaluate the communication negatively. However, the $1.00 experimental condition is assumed to have the effect of imposing a positive unit relation of responsibility between source and communication which is resistant to change since the positive inducement to engage in the discrepant behavior is low.[8] The structure represented in s-digraph a is unbalanced. We assume that there is a tendency for the attitudinal relations to be modified in the direction of those involved in the balanced structure represented in s-digraph b, i.e., a positive attitudinal relation towards the experiment and a positive evaluation of the communication. This type of modification is less likely in the $20.00 experimental condition where the subject may feel less responsible for his action. Here the positive unit relation is likely to be fairly weak since the inducement to send the attitude-discrepant communication is high. Hence the source can view his action as an outcome of the high financial reward offered rather than as expressing his own desire to communicate the position advocated. The positive unit relation is therefore more easily changed in sign. The source can disown the communication and maintain his negative attitude towards the experiment (s-digraph c). Figure 9 also presents a second type of unbalanced structure (s-digraph d) and the corresponding balanced structures which are likely to develop if the positive unit relation is very strong (s-digraph e) or very weak (s-digraph f). The above analysis assumes that the position taken by the communication is quite unambiguous and cannot be misinterpreted by the source.

The preceding discussion suggests that it may be possible to interpret many of the dissonance studies involving attitude–discrepant behavior in terms of the degree to which the person considers himself responsible for the behavior. Brehm and Cohen (1962), for example, summarize studies which imply that the magnitude of dissonance arousal tends to be greater the less the positive inducement for commitment to discrepant behavior, the greater the choice in commitment to discrepant behavior, and the less the coercion applied in order to induce discrepant commitment. We believe that each of these experimental operations may influence the degree to which the person believes that he is personally responsible for his actions. Conditions of high positive inducement, low choice, and high coercion would function to reduce the tendency for the person to feel that he is responsible for his attitude–discrepant behavior. These are factors introduced by others into the situation. A person can appeal to these factors to justify behavior which is discrepant with his attitude. Where positive inducement is low, degree of choice is high, and coercion is minimized however, the person is less able to justify attitude-discrepant behavior in terms of external constraints. Under these conditions he is more responsible for his behavior. Brehm and Cohen (1962) express a similar idea when, in discussing experimental evidence, they conclude that, "The greater the degree of volition involved in the occurrence of dissonant cognitions, the greater the magnitude of the dissonance [p. 217]." The above analysis of Festinger and Carlsmith's experiment suggests that within the structural balance model "volition" or responsibility may be conceptualized by a positive unit relation.

---

[8] It is likely that some subjects would disown responsibility for the attitude–discrepant behavior even in the $1.00 experimental condition. They could decide, for example, that they actually had little choice in the matter since the experimenter would expect their cooperation. They therefore acted for him rather than of their own volition. It would in fact be possible to incorporate the experimenter as another element within the structural balance model and to represent relations between the experimenter and S and the experimenter and C (e.g., the experimenter is seen by S as a trustworthy authority who is obviously in favor of C). In some cases S may justify attitude–discrepant behavior because a positively regarded experimenter is in favor of the behavior. This might influence S's initial attitude towards the behavior as well as his perceived responsibility for the behavior.

CONCLUSIONS

The preceding discussion of experimental studies shows that the structural balance model has considerable power in providing the possibility of a general conceptualization of communication effects. We have shown that one may interpret studies based on different theoretical approaches in terms of the one general model.

In using this model we map into the cognitive space of either the source or the receiver of the communication. If, for example, we are interested in the receiver's reactions to a communication then the relations in the communication structure would be considered from his point of view. Similarly, if we are interested in the source's reactions to a communication then the communication structure would be in terms of his perception of the different relations. As Osgood (1960) indicates,

> Cognitive interactions, obviously, transpire within the nervous systems of single individuals. The 'maps' we draw to represent such interactions necessarily reflect person-person and person-object relations *as some individual perceives them* [p. 346].

The studies discussed above imply the feasibility of varying relations within this cognitive structure by experimental operations. It is then possible to observe behavior which is assumed to reflect other relations within this cognitive structure, or changed relations in accordance with the balance principle. For example, suppose that in an experiment a speaker who was introduced as a world authority on a particular issue then delivered a speech deliberately intended to be at variance with the subject's known attitude towards the issue. We assume that this experimental manipulation would initially determine an unbalanced cognitive structure such that the receiver would be positively disposed towards the source, he would perceive the source as responsible for the communication, and he would see the attitude of the communication towards the issue as different from his own. We would predict that this unbalanced cognitive structure would tend to become balanced, and that changes in the relations within the structure would be reflected in the receiver's reactions to the communication. The predictions in this experiment would therefore depend upon our assumptions about the cognitive structure of the receiver since we are interested in his reactions. It should be obvious that in any given communication situation there may be important differences between the cognitive structures of the source and the receiver in that they may not perceive relations within the communication situation in the same way.

There are some important ways in which the structural balance model may be developed further. They are as follows:

1. Further development of the model could consider relations which differ in *strength* as well as in sign. At present, the model considers only relations which are positive or negative. With this development it would be possible to deal more adequately with investigations involving variation in the *discrepancy* between the subject's stand on an issue and the position advocated in a communication about the issue (Hovland, Harvey, & Sherif, 1957; Sherif & Hovland, 1961). One could also provide a more detailed analysis of research findings based upon both the congruity model (Osgood & Tannenbaum, 1955) and dissonance theory (Brehm & Cohen, 1962; Festinger, 1957).

Cartwright and Harary (1956) refer to the development of a systematic treatment of relations of varying strength as one of the principal unsolved problems of linear graph theory. They believe that it is possible to deal with the strength of relations by the concept of a graph of strength σ suggested by Harary and Norman (1958). Morrissette (1958) assumes that the strength of a relation may vary continuously from +1 through 0 to −1. He defines the *strength of a cycle* as the product of the strength of its lines, and the *sign of a cycle* as positive if the product of the strength of its lines is positive, and negative otherwise. Using these concepts he develops a formula for the degree of total balance of a signed graph. The problem goes beyond defining the strength of a cycle or semicycle, however, in that there are many other kinds of questions which one would seek to answer. For example, if the attitudinal relation of receiver to issue is strong (e.g., .9) and the attitudinal relation of communication to issue is moderate (e.g., .5), what will be the strength of the attitudinal relation of receiver to communication[9]? Following Osgood and Tannenbaum (1955) might there be some readjustment in strengths of relations within a semicycle according to a congruity formula? Given semicycles of equal length, is a weak negative semicycle less likely to change than a strong negative semicycle? Given relations of the same strength (e.g., .7), are large unbalanced semicycles less likely to change than short unbalanced semicycles? These kinds of questions can only be answered by further elaboration of the model combined with empirical tests of predictions.

2. The model could also be developed to take account of differences in the *importance* of the issue for either the source or the receiver. Dissonance theory (Festinger, 1957) allows for the importance of the cognitive elements involved in a dissonant relationship (although just how one would determine their "weighting" is rather unspecified), and there have been some recent studies concerned with degree of involvement (e.g., Zimbardo, 1960). Sherif and Hovland (1961) assume that a strong stand on an issue which is personally involving for the individual functions as an anchor in his judgments so that he is less

---

[9] Where the former two relations are of the same sign and both are strong it seems likely that the balanced third relation would be strong. Conversely, where the former two relations are of the same sign and both are weak it seems likely that the balanced third relation would be weak.

likely to be influenced by highly discrepant persuasive communications. This type of variable implies resistance to change in the attitudinal relation of person to issue. It may, however, be an oversimplification to equate the importance of the issue with the strength of the person's attitude towards the issue since attitudes may presumably be of equivalent strength yet differ in their resistance to change. For example, one might feel strongly about jazz music and strongly about scientific integrity but the second attitude, it is hoped, would be less amenable to change than the first. The intensity or strength of an attitude is probably only one factor governing its resistance to change. Katz (1960) suggests that other important factors are the specificity or generality of the attitude, its degree of differentiation, the number and strength of its linkages to a related value system, and the centrality of the attitude. The latter factor refers to the role of the attitude "as part of a value system which is closely related to the individual's self-concept [Katz, 1960, p. 169]." If these different factors are to be taken into account it is apparent that the conceptualization of the importance of the issue in terms of resistance to change in an attitudinal relation presents a formidable task.[10]

3. One aspect of the structural balance model that requires some clarification concerns the *specific effects* of a state of imbalance within a structure. The relations within a structure are assumed to be modified so that the structure becomes balanced. But which relations will be modified? In representing the outcomes of some of the experimental studies considered in the previous section we assumed that the weakest relation in an unbalanced structure will tend to change. In Figure 10, for example, by this assumption the attitudinal relation of receiver to issue represented in s-digraph a would tend to change to positive since it is the weakest relation, but in s-digraph b the weakest relation represented is the attitudinal relation of receiver to source and it would tend to change to negative. Such an analysis would lead one to expect that when a highly credible source is seen by the receiver as obviously responsible for a communication having a strong attitude towards an issue which is at variance with the receiver's attitude towards the issue, then the receiver's attitude may change. Where, under similar circumstances, the source is only moderately credible, the receiver may maintain his attitude towards the issue and disparage the source. Results of a recent study by Aronson, Turner, and Carlsmith (1963) tend to be consistent with this prediction.

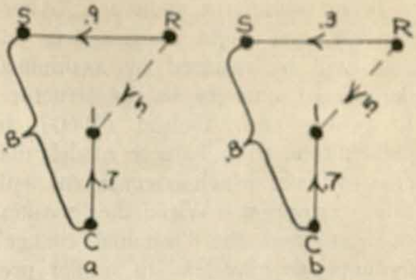


FIG. 10. Unbalanced s-digraphs (a and b) differing in strength of attitudinal relation of receiver to source.

These are, however, other possible assumptions. Following Osgood and Tannenbaum (1955), one might assume that multiple changes will tend to occur within the structure. Alternatively, one might assume that balance will be restored by a minimal number of changes in the structure. Rosenberg and Abelson (1960), in presentation of a balance model, discuss evidence which is consistent with this assumption. While the "weakest relation" and the "minimal change" assumptions may lead to similar predictions for some communication situations, it is possible that there are cases where predictions would differ. It is obvious, however, that different assumptions about specific effects within unbalanced communication structures need to be explored experimentally. Dissonance theory (Brehm & Cohen, 1962; Festinger, 1957) seems to avoid this issue and its related experiments generally attempt to block off all possible avenues of dissonance reduction except the one being studied.

---

[10] One possible approach would be to consider the issue as one of a set of issues, closely bound by unit relations, and towards which the receiver has strong attitudinal relations. In this case a change in any one attitudinal relation might produce imbalance in the whole system and would be resisted.

4. Finally, the structural balance model needs to take account of *individual differences* in the degree to which persons are able to tolerate inconsistencies. One might do this by using the concept of a threshold for intolerance of inconsistency and by specifying the different factors which affect the height of this threshold (cf. Hovland & Rosenberg, 1960). It is likely that this threshold is a function of both personality and situational characteristics. In a recent study of personality variables (Feather, 1964) we investigated the effect of differences in attitude strength, critical ability, and intolerance of ambiguity on the tendency to evaluate logical arguments in a manner consistent with attitude. There is, however, a dearth of studies concerned with individual differences in reactions to imbalance, incongruity, etc., and there is an important need for research in this relatively unexplored area.[11]

It will be evident that the possible developments in the structural balance model listed above are also relevant to most of the other consistency models. These models face similar issues. We believe that these issues can be resolved in the structural balance model by additional theoretical assumptions whose implications are checked empirically. This would lead to a more complex mathematical model with considerable scope. Even now, however, the structural analysis of communication situations outlined in this paper provides something of a unifying approach to the analysis of communication effects. It suggests also that there are many unbalanced communication structures awaiting investigation, since Figure 3 implies that the number of balanced communication structures is quite limited.

The approach discussed in the present paper is at a "molar" level. It specifies the *conditions* for balanced structures but does not attempt to explain just *why* these are the conditions. Some suggestions about underlying mechanisms have already been made in the literature (cf. Hovland & Rosenberg, 1960, pp. 223–227), but much more needs to be done to account for the origins and development of preferences for balanced and consistent structures. The present analysis of structural relations based on the theory of linear graphs does, however, provide a unifying and testable model with many interesting theoretical and research implications.

---

[11] A particular individual's reaction to imbalance would depend upon the resistance to change of relations within his cognitive structure. An individual may also be more intolerant of imbalance in some situations than in others depending, for example, on whether or not his evaluation of the communication is publicly expressed (cf. Hovland & Rosenberg, 1960).

REFERENCES

Aronson, E., Turner, J. A., & Carlsmith, J. M. Communicator credibility and communication discrepancy as determinants of opinion change. *J. abnorm. soc. Psychol.*, 1963, **67**, 31–36.

Brehm, J. W., & Cohen, A. R. *Explorations in cognitive dissonance.* New York: Wiley, 1962.

Brown, R. Models of attitude change. In R. Brown, E. Galanter, E. H. Hess, & G. Mandler (Eds.), *New directions in psychology.* New York: Holt, Rinehart, & Winston, 1962.

Cartwright, D., & Harary, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, 1956, **63**, 277–293.

Feather, N. T. Subjective probability and decision under uncertainty. *Psychol. Rev.*, 1959, **66**, 150–164.

Feather, N. T. Cognitive dissonance, sensitivity, and evaluation. *J. abnorm. soc. Psychol.*, 1963, **66**, 157–163.

Feather, N. T. Acceptance and rejection of arguments in relation to attitude strength, critical ability, and intolerance of inconsistency. *J. abnorm. soc. Psychol.*, 1964, in press.

Festinger, L. *A theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.

Festinger, L., & Carlsmith, J. M. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203–210.

Harary, F., & Norman, R. Z. Graph theory as a mathematical model in social science. Ann Arbor, Mich.: Institute for Social Research, 1953.

Heider, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Hovland, C. I., Janis, I. L., & Kelley, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.

Hovland, C. I., Harvey, O. J., & Sherif, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

Hovland, C. I., & Rosenberg, M. J. (Eds.), *Attitude organization and change.* New Haven: Yale Univer. Press, 1960.

Katz, D. The functional approach to the study of attitudes. *Publ. Opin. Quart.*, 1960, **24**, 163–204.

Kelman, H. C., & Hovland, C. I. "Reinstatement" of the communicator in delayed measurement of opinion change. *J. abnorm. soc. Psychol.*, 1953, **48**, 327–335.

Newcomb, T. M. *The acquaintance process.* New York: Holt, Rinehart, & Winston, 1961.

Newcomb, T. M. Individual systems of orientation. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 3. New York: McGraw-Hill, 1959. Pp. 384–422.

Morrissette, J. An experimental study of the theory of structural balance. *Hum. Relat.*, 1958, **11**, 239–254.

Osgood, C. E. Cognitive dynamics in the conduct of human affairs. *Publ. Opin. Quart.*, 1960, **24**, 341–365.

Osgood, C. E., & Tannenbaum, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.

Rosenberg, M. J., & Abelson, R. P. An analysis of cognitive balancing. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change.* New Haven: Yale Univer. Press, 1960. Pp. 112–163.

Sherif, M., & Hovland, C. I. *Social judgment.* New Haven: Yale Univer. Press, 1961.

Zajonc, R. B. The concepts of balance, congruity, and dissonance. *Publ. Opin. Quart.*, 1960, **24**, 280–296.

Zimbardo, P. G. Involvement and communication discrepancy as determinants of opinion conformity. *J. abnorm. soc. Psychol.*, 1960, **60**, 86–94.

(Received June 17, 1963)

Psychological Review
1964, Vol. 71, No. 4, 314–330

NOISE AND WEBER'S LAW:

THE DISCRIMINATION OF BRIGHTNESS AND OTHER DIMENSIONS

MICHEL TREISMAN[1]

Institute of Experimental Psychology, Oxford

The model of the threshold given by signal detection theory is used as a basis for explaining the form of the Weber function. For this purpose three sources of noise are considered: spontaneous background noise, the irreducible physical variability of the stimulus, and variation in the response of the neural pathways simultaneously excited by the administration of a stimulus. It is postulated that the last will show positive correlation, i.e., that the central effects of the different elements which are simultaneously activated by a stimulus will vary from trial to trial in a similar, rather than in independent fashion. If this is accepted, a law for the Weber function can be derived which gives, for visual brightness discrimination, a square root relation at low stimulus intensities with a continuous transition to a linear law at high intensities, and which can be applied to explain some features of visual spatial and temporal summation. The model also allows Weber's law to be derived for other sensory dimensions.

Weber's law states that $\Delta I/I = k$, where k is a constant, I is a given stimulus intensity, and ΔI is the just noticeable difference in that intensity, according to some criterion (e.g., that it is detected on 50% of trials when it is added to I). But the Weber functions (the relations between ΔI and I) which are determined experimentally for the different modalities are not well described by Weber's law. The law holds, to an approximation, for the midranges of many stimulus dimensions, but for low values of I, and sometimes for high values, the Weber fraction $\Delta I/I$, tends to increase. The linear generalization of Weber's law, given by adding a constant, I_n, to I, usually gives a better fit, as was known to Fechner (1860); and Guilford (1932) has suggested that the power of I should be assumed to vary between .5 and 1. Thus a general expression for the Weber function might be

$$\Delta I = k(I + I_n)^p \qquad [1]$$

where k, p, and I_n are constants.

The Weber function helps to define the input-output relations of human subjects performing discrimination tasks. It limits the logical space within which possible mechanisms for discrimination may be postulated; any such model must give this function as its output. This paper will attempt to show that the model based on signal detection theory can, with additional assumptions, adequately account for the main observed features of Weber functions.

The combination of statistical decision theory (Neyman & Pearson, 1933) and the hypothesis that successive presentations of a single stimulus produce a dispersion of central neural effects (Thurstone, 1927a) led to the development of signal detection theory (Barlow, 1956; de Vries, 1943; Gregory, 1956; Rose, 1948; Swets, Tanner, & Birdsall, 1961; Tanner & Swets, 1954). Aspects of the theory relevant to the arguments to be presented will be shortly described. (See Figure 1). We assume that when stimuli varying on a physical dimension I are presented to a subject they produce central neural effects whose magnitude is described by $E = f(I)$. The dimension E serves as the "decision axis"; each stimulus presentation produces a magnitude of E, and a decision rule is applied to this magnitude to determine which response the subject will make. When a stimulus I' is repeatedly presented it produces a distribution of central effects with mean E' and variance $\sigma_{E'}^2$. In the simplest case it is assumed that this distribution is approximately normal, and that for stimuli I', I'', I''', lying in the small range used to determine a single threshold, $\sigma_{E'} \doteq \sigma_{E''} \doteq \sigma_{E'''} = \sigma_E$. If on each trial of an experiment the subject must decide whether or not the stimulus presented was greater than I', the decision rule would be that he must compare each central effect, E, with a criterion, c_E; if E exceeds this value the response is "Yes," if not, "No." For trials on which a Stimulus $I'' > I'$ is presented the probability of the response "Yes," $P(Y|I'')$, would then be given by the integral of the E'' distribution for $E > c_E$; when I' is presented the corresponding integral of the E' distribution would give the false positive rate, $P(Y|I')$.

---

[1] I would like to thank H. B. Barlow, F. H. C. Marriott, M. G. Bulmer, and P. E. Hallett for reading the manuscript and for very helpful comments and criticisms. I would also like to thank H. B. Barlow for permitting me to reproduce Figure 2. My appreciation is due to R. C. Oldfield for the provision of research facilities.

There are a number of alternative methods for computing the optimal criterion, c_E (Birdsall, 1955); however, the criterion proposed by Neyman and Pearson (1933) has the advantage that it is applicable to the common psychophysical procedures, and to everyday situations (Treisman, 1964a). It requires that the false positive rate must not exceed a limiting value, and is satisfied if $c_E = E' + z\sigma_{E'}$, where z is a constant such that $P(Y|I')$ equals the limiting acceptable false positive rate. This does not require that the subject know the form of the E'' distribution, or the a priori probability of I''.

Figure 1 shows the correspondence between the distributions on E and the psychometric function. We can consider the E' distribution as a "noise" distribution and the E'' distribution as due to "signal + noise." The figure applies to the determination of the incremental threshold or of the absolute threshold by the method of constant stimuli. In the latter case $I' = 0$, in the former the variable stimulus increments, ΔI_v, are superimposed on a constant background intensity, $I' > I_0$, where I_0 is the absolute threshold. In each case the threshold criterion, $P(Y) = .5$ (where $P(Y)$ is the probability of the response "Yes"), will be satisfied by the stimulus $I'' = I' + \Delta I$, for which $E'' = c_E = f(I'')$. When $I' = 0$, I'' is the absolute threshold; when I' is supraliminal ΔI is the just noticeable difference, and corresponds to $z\sigma_E$. The standard deviation of the psychometric function, σ_p, will correspond approximately to $I''' - I''$, where I''' is the stimulus for which $P(Y) = .84$; the central effect of I''' is $E''' = E'' + \sigma_{E'''}$.

Another procedure sometimes used to measure the just noticeable difference is to present the standard stimulus I' and a variable stimulus $(I' \pm \Delta I_v)$, requiring the subject to report whether the second is greater or less than the first. In this case the psychometric function for

"Greater" responses will have its 50% point at I' (ignoring any constant error) and the standard deviation of this psychometric function may be taken as the differential threshold (this corresponds to an 84% response criterion). If on each occasion the decision device compares the central effect of $(I' \pm \Delta I_v)$ with E', then the standard deviation of the

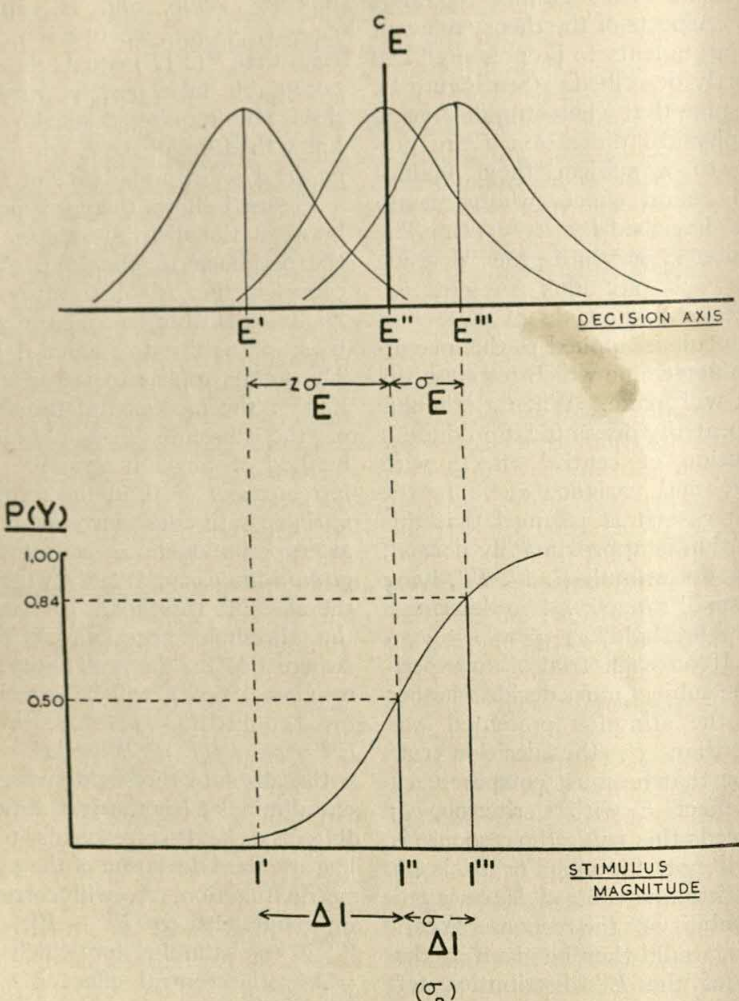


FIG. 1. The signal detection model for the measurement of the absolute or incremental threshold by the method of constant stimuli. (Repeated presentation of a stimulus I produces a dispersion of central effects on the dimension E. This dimension serves as the "decision axis" used to determine the responses when threshold discriminations are made. The psychometric function for $P(Y)$, the probability of the response "Yes" is shown. The probability of the response "Yes," $P(Y)$, corresponding to I', the standard stimulus, is the false positive rate. I'' is the stimulus giving $P(Y) = .50$; $I'' - I' = \Delta I$. $I'' + \sigma_p = I'''$, where σ_p is the standard deviation of the psychometric function. With these procedures, σ_p is identical with $\sigma_{\Delta I}$, the standard deviation of ΔI.)

psychometric function will again correspond to σ_E. But if the procedure is such that the subject cannot identify the standard stimulus as such on each trial, and he must operate by comparing the particular central effects produced by each of the two stimuli on each trial, then, if these are uncorrelated, the jnd will correspond to $\sqrt{2}\sigma_E$ (Thurstone, 1927b). Thus with each method the jnd will be a function of σ_E. The value of ΔI given by a procedure of this sort is sometimes referred to as the "differential threshold." However, in what follows we shall consider mainly the incremental threshold.

Weber Function

The signal detection model for threshold discrimination has now been presented, and we must examine whether it can be used to explain the Weber function. It follows from the model that the relation between ΔI and I is to be understood in terms of the functions relating E to I, and σ_E to E. Thus the sources of the noise which produces the dispersion of the central effects of a nominally constant stimulus, and contributes to its mean effect, must be analysed. Three sources of noise can be distinguished; in some respects they can be best differentiated in relation to visual intensity discrimination, and this modality will be mainly discussed here.

The importance of spontaneous activity in the sensory system in the absence of a stimulus was first recognized by Fechner (1860) who attributed departures from Weber's law at low visual intensities to the "Augenschwarz." Kuffler, FitzHugh, and Barlow (1957) observed maintained discharges in the ganglion cells of the cat in complete darkness, and they suggested these might arise from the decomposition of visual pigment, excitation of fine dendritic terminals by random electrical noise, or in other ways. Noise due to spontaneous firing, or to the incursion into the visual pathways of impulses originating outside them independently of the stimulus, might also arise at various levels of the afferent system. For simplicity, however, this background noise will be treated, following Barlow (1957a), as though it is caused mainly by thermal decomposition of molecules of visual pigment. The spontaneous background activity is equivalent in effect to a stimulus of small intensity which is constantly present, or "dark light"; it explains the existence of an absolute threshold, and is represented by the constant I_n in Equation 1. Thus it accounts for the improved fit when the linear generalization of Weber's law is used instead of the classic form. The background activity will mainly affect the mean of the central distribution corresponding to a given stimulus; it does not account for the constants k and p.

A second important source of noise is the irreducible physical variability of a stimulus of constant nominal intensity. With light, this is the variation in the number of quanta absorbed from a flash of constant intensity (or considered to be absorbed from the dark light), which is described by the Poisson distribution. Rose (1948) compared the eye to an ideal device whose performance is limited only by fluctuations in the absorption of quanta, and to this Barlow (1958a) added the assumption of a dark current. Thus if the mean number of quanta, m, effectively absorbed from a flash of constant intensity, I, is given by $m = gI$, where g is a constant, I_n is the dark light, and detection is limited only by the variance of the number of quanta absorbed, then the incremental threshold, ΔI, will be $zg^{-\frac{1}{2}}(I + I_n)^{\frac{1}{2}}$. Therefore the model based on these two noise sources predicts a power function with the exponent $p = .5$.

Agreement with the square root law has been found for stimuli of small area and short duration on backgrounds of low intensity for rod vision (Barlow, 1957a) and, under similar conditions, for cone vision (Barlow, 1958a). But under other conditions, "the experimental points deviate from the appropriate theoretical curve and tend to obey the Weber law instead [Barlow, 1958b]." Since most studies have found good approximations to the linear generalization of Weber's law for the greater part of the range of stimulation (Aguilar & Stiles, 1954; Blackwell, 1946; Steinhardt, 1936), we face the problem "can reasons be found for the departures from the theoretical predictions which occur when the test stimulus is of long duration or large area, and when the background intensity is high [Barlow, 1957a]?"

An approximately linear Weber function is so ubiquitous in studies of sensory discrimination that any explanation for it for visual intensity should be capable of generalizing to at least some other modalities. Since, on the signal detection model, noise will limit threshold discrimination on all dimensions, and the two sources of noise considered above only partly explain the form of the Weber function, it is prima facie plausible that the occurrence of linear Weber functions may be due to a feature of the third source of noise which is not specific to vision alone. This source is the "neural noise" which may arise from variation in the responsiveness of the afferent paths to the sensory inputs resulting from the absorption of quanta from the stimulus (or from the dark light). Since little is known about this noise, assumptions must be made if predictions are to be derived from it. Two hypotheses will be presented and their consequences discussed; evidence for them will be considered later. They are: (*a*) The effective absorption of a quantum of light (or a noise event equivalent in its effects to the breakdown of a molecule of visual pigment) will produce a sensory message of mean magnitude s and with variance σ_s^2. If m quanta are effectively absorbed from a light flash, and n equivalent noise events occur, the central effect of that flash will be given by $E = \sum_{i=1}^{m+n} s_i$. For flashes of constant nominal intensity I the mean central effect will be

$$E = sg(I + I_n) \qquad [2]$$

This hypothesis is discussed and modified below. (*b*) One way in which the afferent visual system may be conceived is as a set of parallel afferent paths, each of which may serve to convey the sensory message resulting from the absorption of a single quantum of light. The total central effect resulting from a given stimulus administration is then the sum of the messages carried by all the paths activated by that stimulus. The responsiveness of each path varies from trial to trial, so that the central effect produced by the absorption of a quantum by a given receptor on different trials will vary; this variation in responsiveness is the source of σ_s^2. Furthermore, the different paths do not vary independently in their responsiveness, but in a related way, so that on one trial the responsiveness of all paths may be greater than usual, on another trial less. These assumptions about the afferent systems are quoted to illustrate a possible physiological basis, not necessarily the only one, for the second hypothesis: that the product-moment correlations r_{ij} between s_i and s_j, where these are any two of the sensory messages produced by the absorption of quanta from a given light flash (or by equivalent noise events), will be positive; we shall suppose them to approximate to unity. Since s and $g(I + I_n)$ are independent, and the latter is described by the Poisson distribution, it will then follow from Equation 2 that:

$$\sigma_E{}^2 = \sigma_s{}^2 g^2 (I+I_n)^2 + s^2 g(I+I_n) + \sigma_s{}^2 g(I+I_n) \qquad [3]$$

and consequently:

$$\Delta I = \frac{z}{sg}\,\sigma E = \frac{z}{g}\sqrt{g^2 v^2 (I+I_n)^2 + g(1+v^2)(I+I_n)} \qquad [4]$$

where $v = \sigma_s/s$ is the "coefficient of variation" of the sensory message.

Since z, g, v, and I_n are constants, it can be seen that if $v^2 \ll 1$, which is plausible, then for $g(I + I_n)$ small the function will approximate

$$\Delta I = zg^{-\frac{1}{2}}(I + I_n)^{\frac{1}{2}} \qquad [5]$$

and as $g(I + I_n)$ becomes large there will be a transition to a function approximating

$$\Delta I = zv(I + I_n) \qquad [6]$$

Thus the assumptions made lead immediately to the transition from a square root law at low intensities to the linear Weber function at high intensities which required explanation. We may note that the exponent p in Equation 1 can now be interpreted as a measure of the extent to which positive correlation between the sensory messages on each trial determines the variance of E, and k can be defined in terms of z, g, and v.

If this model correctly represents the phenomena which underlie the Weber functions, it will appear that Nature has been unkind to the experimental psychologist in burdening him with a relation dependent on four parameters. However this number does not present an insuperable bar to devising experimental tests of features of the model, as can be seen if the parameters are considered individually.

1. I_n. Barlow (1957a) estimated I_n by plotting $\log \Delta I$ against $\log I$ and taking the intensity of I at which the curve fitted to these values intersected the ordinate corresponding to the absolute threshold as the dark light; Gregory (1956) estimated I_n by extrapolating the linear portion of the curve relating ΔI to I to an intercept with the abscissa. The estimates obtained probably depended on rod mechanisms in the former case and cones in the latter; but if a Weber function were determined with precautions taken to make it depend on the rods alone, then independent estimates of I_n could be obtained from the portion of the curve described by Equation 5 and that described by Equation 6, and we could examine whether they are the same, as, on the present assumptions, they should be. Another procedure for estimating I_n from absolute threshold judgments, which, however, is not very exact, is given by Barlow (1956).

2. z. This parameter can be eliminated from Equation 4, or it can be estimated in two independent ways. (*a*) If instead of the incremental threshold we measure the differential threshold by the method of constant stimuli, requiring the responses "Greater" or "Less," the resulting psychometric function will have its

50% point at I', the standard stimulus (ignoring any constant error). If we take the standard deviation of the psychometric function, σ_p, as ΔI, this will correspond to σ_E, not to $z\sigma_E$, so that z will disappear from Equations 4, 5, and 6. Thus this psychophysical procedure would give a Weber function dependent on only three parameters. (*b*) Crozier's law (Crozier, 1936; Crozier & Holway, 1937; Holway, 1937; Holway & Crozier, 1937; Upton & Crozier, 1936) can be written $\Delta I/\sigma_{\Delta I} = C$, where C is constant and $\sigma_{\Delta I}$ is the standard deviation of ΔI (i.e., σ_p when the method of constant stimuli is used). This relation could be written $(I'' - I')/(I''' - I'') = C$ (see Figure 1), and this corresponds to $(E'' - E')/(E''' - E'') = z\sigma_E/\sigma_E = z$. When E is linearly related to I (for the ranges used in determining thresholds) $C = z$ exactly, and even if the relation is nonlinear it appears that, in some cases, C may still not differ much from z (Treisman, 1964c). Thus, if the incremental threshold is measured, C may be calculated and taken as an estimate of z. (*c*) If the false positive rate, $P(Y|I')$, is recorded for the incremental threshold, z can be determined by reference to a table of the normal curve.

It should be possible to estimate z by both procedures, and compare the two results obtained (though the large number of observations required to determine a small false positive rate accurately is a problem in practice), or to use these estimates when attempting to fit Equations 5 or 6 to experimental data.

3. v. If z is known v could be obtained from the slope of the linear portion of the Weber function. Since $v = \sigma_s/s$, any factor which would tend to increase the variability of the neural response would be expected to increase the slope of the linear portion of the Weber function. Age might be thought to be such a factor, a prediction made and tested by Cane and Gregory (1957), who found that the slope, but not the intercept with the abscissa, of the linear portion of the curve increased in older subjects; they interpreted this as indicating an increase in neural noise but not in retinal noise (I_n).

4. g. This constant relates the number of quanta effectively absorbed to the intensity of the stimulus. It will depend on the area and duration of the stimulus, and on f, the fraction of the quanta incident at the cornea which is effectively absorbed by the visual pigment. The value of f can be estimated by consideration of the physical and chemical properties of the eye and retina; Rushton (1956) estimates that 10% of light entering the eye ($\lambda = 505$ mμ, 20° parafoveal) is absorbed by rhodopsin. Thus estimates of g, derived in this way, z and I_n could be substituted in Equation 5 to predict the exact form of the lower portion of the Weber function; or a value of g could be derived from the threshold data, applying Equation 5, and compared with that given by physical considerations.

A comparison of Equations 5 and 6 will suggest another prediction: since f, and therefore g, is a function of wave length, and g is present in the former equation but not the latter, wave length should affect the Weber function for low but not for high stimulus intensities. This expectation is supported by observation: "at high intensities the Weber fraction $\Delta I/I$ of the cone system as a whole . . . becomes constant (Weber's law) and has about the same value for light of all spectral compositions [Pirenne, 1962]."

Since g depends on the area and

duration of the stimulus, a further test of the model would be given by seeing whether predictions that can be derived from Equation 4 for variation in the area and duration of the stimulus agree with experimental findings on spatial and temporal summation. This question will be examined more closely in a later section, but first further consideration will be given to the two assumptions made in deriving the model.

Visual Pathways

Relatively little is known of the detailed functioning of the afferent pathways; nevertheless, if the two assumptions on which Equation 4 is based are to be accepted they must be compatible with what is known of sensory physiology. They are, of course, simplifications which need not hold exactly at the physiological level. Linearity need not be perfect nor correlation complete for Equation 4 to fit the data to a good approximation. In this section some of the possible ways in which neural mechanisms underlying vision might function will be discussed, mainly in order to demonstrate that the hypotheses given above could describe the behavior of physiological bases which are not completely implausible, and without wishing to imply that other physiological assumptions could not be made which would be equally or more acceptable.

We start by supposing that the effective absorption of a quantum of light causes a train of impulses to be transmitted from the receptors, through bipolar cells, to the ganglion cells, where the different trains set up by the absorption of quanta from a given flash sum to give E, the linear relation given in Equation 2 applying to this transmission. The neural noise arises from variation in transmission at the retinal synapses, i.e., variation in the number of impulses produced in a bipolar cell when a quantum of light is absorbed by its receptor, or in the number of postsynaptic impulses produced in a ganglion cell by the arrival of a constant presynaptic train. This variation in transmission is the result of the presynaptic train being added to a pre-existing level of facilitation at the synapse, this level varying from trial to trial, and the sum of the two effects determining the postsynaptic response. The facilitation might be due to a bombardment of all the synapses, produced by a single central source whose output varies, and mediated by centrifugal fibers (Granit, 1955). When the output of this source is high transmission will be good at all the synapses, when it is low each quantum absorbed will make a smaller contribution to the central effect than usual, so that the central effects of the different quanta absorbed from any one flash will not be independent but will show some degree of positive correlation. Provided that the greater proportion of the neural noise is assumed to arise at this linear stage in the retina it would not be necessary to require the ganglion cell response also to be a linear function of I. If, for example, the correlated noise occurs before a logarithmic transformation, application of the Neyman-Pearson criterion to the latter would have the same consequences as if it were applied directly at the linear stage, and Equations 5 and 6 would follow as before.

Some evidence on the transmission of the afferent message from the receptor to the ganglion cell is provided by neurophysiological observations which show that Riccò's law holds for single receptive fields of the frog and cat, when the effect of light is measured by its ability to evoke a threshold or constant suprathreshold discharge from the ganglion cell (Barlow, 1953; Barlow, FitzHugh, & Kuffler, 1957; Hartline, 1940). Within limits, $Q = AI$ is constant as A, the area of the stimulus, varies, which is taken to indicate that the effects on the receptors sum at the ganglion cell. If the mean quantity of light, Q, applied to a receptive field is constant, and on any occasion it is distributed over r receptors, where r may vary from trial to trial, then each receptor will receive a quantity of light Q/r, and will transmit a message to the ganglion cell which is a function of this quantity, $f(Q/r)$. The total effect produced at the ganglion cell by a given presentation of the stimulus will then be $r \times f(Q/r)$. This expression will only be constant as r varies if f represents multiplication by a constant, therefore, since Riccò's law holds for the ganglion cell response at levels of illumination at which each receptor is likely to absorb more than one quantum, this indicates that the transmission from the receptors to the ganglion cell must be effectively linear.

A difficulty which would arise if this model were accepted, with the discharge from the ganglion cell taken to be logarithmically related to stimulus intensity (or to have some other nonlinear monotonic relation to it), is that since $\log \sum s \neq \sum \log s$, the model would then be adequate only for stimuli whose area falls within one receptive field, as stimuli larger than this would require the logarithmic outputs from two or more fields to be summed centrally; consequently, it would become difficult to explain the similar laws for spatial and temporal summation (see the next section) when large areas are used. (The same difficulty arises if an attempt is made to attribute Weber's law to the addition of constant variance after a logarithmic transformation.) But it is not plausible, on the other hand, to suppose that Equation 2 applies exactly to the ganglion cell discharge over the whole range of light intensities. Fortunately, it is not necessary to make so strong an assumption. Since in determining visual thresholds the eye is usually adapted to a fixed intensity of I, to which near-threshold increments, or decrements, are added, we need only postulate approximate linearity for near-threshold ranges in the adapted condition. Thus the effect of adaptation to a constant stimulus, I, might be to eliminate the response to a constant proportion of the mean number of quanta absorbed from it, $hg(I + I_n)$, $h < 1$ being constant, so that only quanta absorbed in excess of this effective zero point on any trial would produce a neural message, with the variance of $g(I + I_n)$, and near-threshold increments or decrements in I, producing approximately linear changes in this message. On this assumption Equation 5 would be unchanged, and Equation 6 would become $\Delta I = zv(1 - h)(I + I_n)$, which is again a linear function. Or we might suppose that when the eye is adapted to I a neural message, given by $s \log hg(I + I_n)$, is produced, to which any excess of stimulation adds a linear increment. On this assumption Equation 4 would become

$$\Delta I = \frac{z}{g}\sqrt{g(1 + v^2)(I + I_n) + v^2[g(1 - h)(I + I_n) + \log hg(I + I_n)]^2}$$

and this again would give Equation 5 for small intensities of $g(I + I_n)$; as intensity increases $g^2(I + I_n)^2$ would grow more rapidly than $g(I + I_n)$ or $\log g(I + I_n)$, with the result that at high stimulus intensities the term giving a linear function would again be the most important in the equation.

On these assumptions the variance due to neural noise could arise at the retina or at higher synaptic junctions, being caused by a single source of facilitation which might be identified with "reticular arousal," or in some other way.

The assumption that small changes in intensity about a level to which the eye is adapted will cause approximately linear changes in the ganglion cell discharge receives some support from observations made by FitzHugh (1957). He examined the effect of brief flashes of light of near-threshold intensity on the discharge from the cat's ganglion cell, recording the relation between the number of impulses produced in a critical period following a flash and the intensity of the latter; he fitted power functions to the data for individual cells and obtained mean exponents of .6–1.2. There is also some psychophysical evidence tending to support him (Treisman, 1964c).

There is little evidence to indicate whether or not sensory noise is correlated. However, an ingenious experiment by Diercks and Jeffress (1962) is of interest. When a tonal signal is presented in white noise the threshold for the tone is lower the greater the disparity between the interaural correlations for the tone and for the white noise (Hirsh, 1948; Jeffress, Blodgett, Sandel, & Wood, 1956). Thus if the tone and noise are both in phase at the two ears, the threshold will fall if the tone at the input to one ear is removed, and will fall further if it is returned to that ear but 180° out of phase with the tone on the other ear. Diercks and Jeffress measured the absolute threshold for a 250 cps tone given alone, and found that the threshold was higher when the signals to the two ears were in phase than when they were 180° out of phase, which they interpret as indicating that the signal is added to internally generated noise with some degree of positive interaural correlation.

A square root law was first suggested by Cattell (1893), who argued,

> The algebraic sum of a number of variable errors tends to increase as the square root of the number. In measuring the base line of a survey the variable error of observation increases as the square root of the length of the line. It seems to me the same relation might be expected to hold in a general way when the length of a line is estimated by the eye. . . . I should substitute for Weber's law the following: *The error of observation tends to increase as the square root of the magnitude.*

Fullerton first suggested that correlation between "fractional or elemental" errors of observation might contribute to the appearance of Weber's law (Cattell, 1893); later Woodworth (1914) argued that "The reaction of the organism to a stimulus is always a complex activity, composed of many elementary activities," and that these might be correlated. However, perhaps in part because of a fallacious rebuttal by Guilford (1932), this suggestion dropped out of sight.

Spatial and Temporal Summation in Vision

The energy in a light flash can be varied by altering its intensity, area or duration. Any account of the effect of change in I on ΔI should, therefore, be capable of being extended to cover spatial and temporal effects in similar detail.

The main features of spatial and temporal summation are similar. For

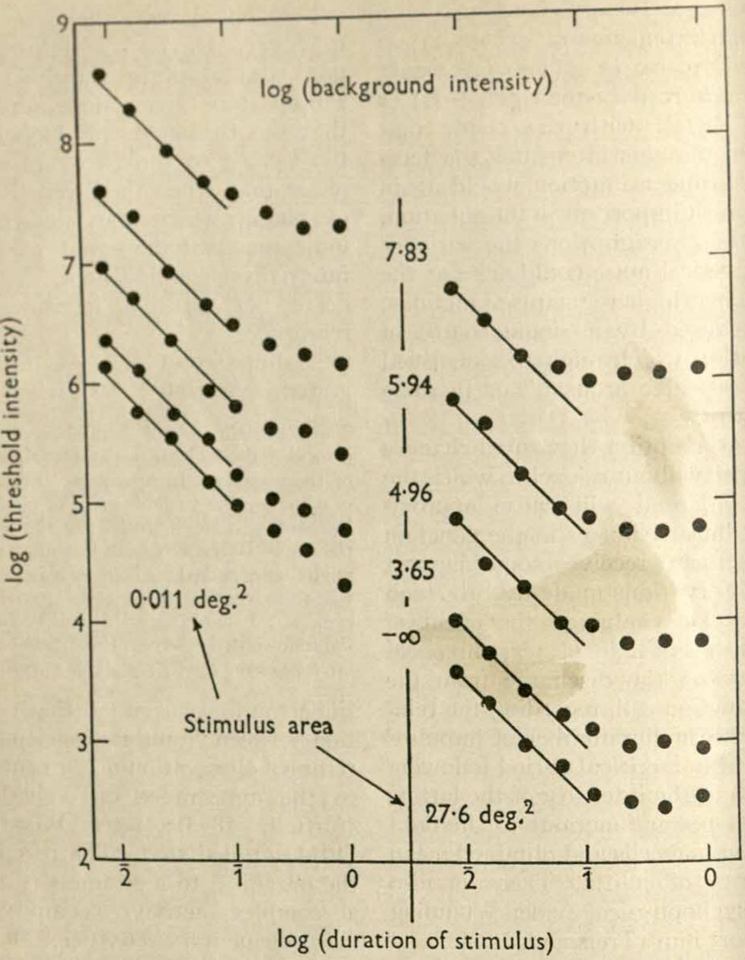


FIG. 2. Data obtained by Barlow (1958b). (Log [increment threshold intensity] is plotted against log [duration] for a small [.011 degree^2] and a large [27.6 degree^2] area of stimulus at five different background intensities. The straight lines have slope −1 and are continued up to .1 second. Intensities are given in quanta [507 mμ]/second degree^2, durations in seconds.)

sufficiently small areas the relation of the absolute threshold, I_0, to the area of the stimulus, A, is given by $I_0A^1 = c$ (Riccò's law). There is then a transition range of area for which Piper's law, $I_0A^{\frac{1}{2}} = c'$, is sometimes claimed to hold, and for large areas the threshold is independent of area, i.e., $I_0A^0 = c''$ (c, c', and c'' being constants). These three relations are described as complete or partial summation, or the absence of summation, the degree of summation being indicated by the exponent of A; this exponent can be represented as the slope of the curve relating log I_0 to log A. Similarly, when the duration, t, of the stimulus is varied, complete temporal summation is obtained for short durations, i.e., Bloch's law (the Bunsen-Roscoe law), $I_0t^1 = k$, holds. Partial temporal summation is obtained in a transition range, and for long durations $I_0t^0 = k''$ (k and k'' being constants). Change in area or duration has similar effects on the difference threshold (Barlow, 1958b; Brindley, 1960; Le Grand, 1957).

Evidence supporting Piper's law or the corresponding square root law for time has been obtained by a number of authors (Baumgardt, 1959; Bouman & van der Velden, 1947; van der Velden, 1944; Willmer, 1950). However, deviations from these laws are frequent (Brindley, 1960; Le Grand, 1957); they are well illustrated by Barlow's (1958b) data (see Figure 2). He measured absolute and incremental thresholds varying both t and I, the background intensity, for different values of A, or varying A and I for different values of t. His results show that, for temporal summation, increase in either the background intensity or the area of the stimulus will tend to reduce the duration at which complete summation ends and the transition range begins, and the duration at which partial summation ends, the transition range becoming shorter, and the mean exponent of t in this range decreasing. The effects of increase in t or in I on spatial summation were very similar. The form of the transition ranges appears to accord with Graham, Brown, and Mote's (1939) finding that the exponent is not constant in this range but decreases "as one progresses from the smallest area to the largest area." Thus it appears that increase in I, A, or t all tend to reduce the exponent of A or t. This gradual change in exponent is quite compatible with the approximate adherence to a square root law over part of the transition range sometimes reported.

Brindley (1960) argues,

> If the relevant noise is assumed to be proportional to the area and duration of the stimulus, as is reasonable for large and long stimuli, the threshold should be inversely proportional to the square roots of area and duration (Piper's law and its temporal counterpart). . . . Since [these laws] are not exactly true, the hypothesis of optimal discrimination of signal from noise cannot be exactly correct.

The form of the signal detection model he thus questions is that based only on background and physical noise. This model also cannot explain why, for still larger areas or longer durations, the threshold becomes independent of area or duration. This makes it of interest to see whether the model developed here, which takes account of neural noise as well as the other sources, can account for summation more adequately, and thus avoid Brindley's stricture.

If, on any trial, an increment ΔI_v is added to a background stimulation of constant intensity I', whether or not it is seen will depend on whether the value of E produced by $(I' + \Delta I_v)$ exceeds $c_E = E' + z\sigma_{E'}$. The values of E' and $\sigma_{E'}$ depend on the area and duration of the background stimulation which the decision device samples. It is often assumed that there is an area α and a duration τ below which complete summation occurs; in terms of the present model these would be the minimum area and duration sampled in reaching a threshold decision. If the area of the increment is A'' and its duration t'', it is simplest to assume that the area and duration sampled on each trial will be $A' = \alpha$ or A'', whichever is the larger, and $t' = \tau$ or t'', whichever is the longer. This implies that g must be separately defined for I' and ΔI_v; g relates the number of quanta which are effectively absorbed and contribute to the decision process to the intensity of the stimulus, and it will depend on the area and duration contributing quanta and on f, the fraction of the quanta incident at the cornea which is effectively absorbed by the visual pigment (Pirenne, 1956; Rushton, 1956). For I' (in quanta/second degree^2 entering the eye), g will be given by $fA't'$, and for ΔI_v by $fA''t''$. We can now rewrite Equation 4 as

$$\Delta I = \frac{z}{fA''t''}\sqrt{f^2A'^2t'^2v^2(I' + I_n)^2 + fA't'(1 + v^2)(I' + I_n)} \qquad [7]$$

For I' and t'' constant, and $A'' < \alpha$ variable, the expression under the square root sign can be written as a constant, c, giving

$$\Delta I = \frac{zc}{fA''t''} \qquad [8]$$

which is, of course, Riccò's law. For $A'' > \alpha$ Equation 7 can be simplified in the same way as Equation 4, giving:

$$\Delta I = \frac{zt'^{\frac{1}{2}}}{f^{\frac{1}{2}}A''^{\frac{1}{2}}t''}(I' + I_n)^{\frac{1}{2}} \qquad [9]$$

for $fA't'(I' + I_n)$ small, and:

$$\Delta I = \frac{zvt'}{t''}(I + I_n) \qquad [10]$$

for $fA't'(I' + I_n)$ large. Equation 9 is Piper's law, and Equation 10 makes the threshold independent of area. Thus as $fA't'(I' + I_n)$ increases there will be a continuous transition between these two laws. An exactly similar argument applies to temporal summation.

Thus the present model accounts for the inadequacy of the square root laws as descriptions of partial spatial and temporal summation, the decrease of the mean exponent in, and the extent of, the transition range as A'', t'', or I' increase and the decrease in the exponent as we go from smaller to larger areas or durations in the transition range. It explains why the duration at which the temporal transition range ends is shorter when A'' or I' are large, and the area at which the spatial range ends is smaller for t'' or I' large, and it accounts for the independence of area or duration found when the latter are very large.

A number of authors have found that increase in A'' will reduce the amount of temporal summation, increase in t'' will reduce the amount of spatial summation, and an increase in I' will reduce both spatial and temporal summation (see Barlow, 1958b). The results in Figure 2 accord with this; Barlow (1958b) notes that the square root laws are never simultaneously true for all three variables—to obtain the square root law for intensity, for example, areas and times within their ranges of complete summation must be used. This agrees with the implications of Equation 7, since the exponent obtained depends on the magnitude of the product $A't'(I' + I_n)$; this product is more likely to be small and so to give a small exponent if the variables held constant are small than if they are large.

If α can vary from time to time, this would explain the occasional exponent smaller than 1 but larger than .5 (Brindley, 1960). It would give a range of areas governed sometimes by Equation 8 and sometimes by Equation 9 and having, therefore, an

[2] E. R. Crossman, personal communication, 1963.

intermediate exponent. That α and τ are not invariable is suggested by the decrease in the area of total summation as t'' or I' increase, or in the duration of total summation as A'' or I' increase (Barlow, 1958b).

> The extent of spatial summation in the living human eye is in broad agreement with the size of the retinal area covered by the ramifications of the large ganglion cells which presumably are responsible for this summation [Pirenne, 1956],

and Barlow, FitzHugh, and Kuffler (1957) have shown that lateral inhibition is reduced in the cat's retina in dark adaptation. If the retina can be considered to consist of "sensitive units" (Brindley, 1960; van der Velden, 1944), i.e., receptive fields summating input over a period of time, it is possible that mechanisms might exist tending to prevent the output of a unit, $f\alpha\tau(I' + I_n)$, exceeding some upper limit, so that if A'', t'', or I' increase there will be "feedback" tending to reduce α and τ. In the case of α this might operate through increase in lateral inhibition.

The threshold intensity is independent of the area and duration of the stimulus when the latter are large, and this is sometimes taken to mean that there is an upper limit of area or time beyond which summation ceases; the threshold model based only on physical and background noise requires such an assumption to explain why the square root law does not apply to large areas and durations. Presumably the stimulus information beyond these limits is supposed not to contribute to threshold decisions; thus the smaller the limit the less efficient the eye. Barlow (1958b) believes his results indicate:

> that big changes in the amount of temporal and spatial summation occur when the background intensity is changed, with the result that when thresholds are determined with a long duration large area test stimulus the experimental points deviate from the appropriate theoretical curve and tend to obey the Weber law instead.

Here the term summation is used descriptively to refer to the occurrence of a fall in threshold intensity as area or duration increases. The present arguments show that we need not attribute variation in summation, in this sense, to an equivalent variation in the action of an underlying summating mechanism with built-in upper limits. On the present model there is no upper limit to the area or duration of the stimulus effectively utilized, and variations in such limits are not needed to account for the different Weber functions obtained when conditions vary. Information is always drawn from the whole area and duration of the stimulus, and threshold discrimination is limited only by the variability of the processes involved, as indicated by Equation 7.

Other Dimensions

In the study of sensory function approximations to Weber's law are ubiquitous, so that an explanation of the function will appear more plausible if it is not peculiar to a particular modality but is fairly general in its application. The present assumptions satisfy this requirement fairly well. As an example we can consider the visual discrimination of the length of lines. We suppose that the presentation of the lines is sufficiently well controlled for variation in the visual angle subtended by a given line to be negligible. When a line of length L is presented, it excites visual "line-detecting elements" whose number is given by $d = kL$, where k is a constant. In the absence of L, spontaneous activity which can be confused with the effect of the line may occur, exciting kL_n line-detecting elements. An active element produces a sensory message s, with variance σ_s^2; the central effect produced by a presentation of L will therefore be $E = \sum_{i=1}^{k(L+L_n)} s_i$. If all the intercorrelations, r_{ij}, between the sensory effects produced each time the line is presented approximate to unity, then the variance of E will be $\sigma_E^2 = k^2\sigma_s^2(L + L_n)^2$, and consequently the Weber function will be $\Delta L = zv(L + L_n)$, where $v = \sigma_s/s$, which is the linear generalization of Weber's law.

The Weber function for temporal discrimination has been derived in a similar way (Treisman, 1963a). The linearity and correlation assumptions are so general that they could be applied to predicting the variance of nonsensory data. Crossman[2] has noted that the standard deviation of reaction times tends to be a constant proportion of the mean, between 10 and 25%, a relation analogous to Weber's law. We might suppose that the length of a reaction time is mainly due to the occurrence of a greater or lesser number of elementary neural events of similar duration, such, perhaps as synaptic crossings, which intervene between the stimulus and response, so that the reaction time is linearly related to this number. The duration of the elementary events which sum to give the reaction time varies in highly correlated fashion from trial to trial: this will lead to "Crossman's law" in the same way that Weber's law was obtained above. This explanation need not conflict with the hypothesis that reaction time is determined by the rate at which information can be processed (Hick, 1952), or the time necessary to execute a sequential decision procedure (Stone, 1960); we could suppose that these procedures are performed by means of the "elementary neural events" postulated above and determine the number necessary, and that variation in the duration of these events is the major source of response variance. However, it is of interest that Stein[3] applies the negative binomial distribution to reaction times and finds that under certain conditions the standard deviation will approximate a constant proportion of the mean.

The account of the Weber function that has been given is necessarily a simplification, since it is unlikely that in actual cases the assumptions made would be wholly met. It has the advantage of being sufficiently general to apply to the Weber function for any sensory dimension, rather than describing a special case only. It is to be hoped that relating Weber's law to the foundation provided by the signal detection model of the threshold may serve as a basis for clarifying its implications for the troubled field of sensory scaling (Treisman, 1961, 1962, 1963b, 1964b, 1964c).

SUMMARY

The signal detection model for threshold discrimination is outlined and applied to some familiar procedures for measuring thresholds. An attempt is made to use this model to explain the form of the Weber function. Three sources of noise limiting discrimination are discussed: the irreducible physical variability of the stimulus; the spontaneous "background" noise to which the stimulus can be considered to be added; and neural noise arising from variation in the response of the pathways transmitting the sensory message centrally. Two assumptions are added to the signal detection model: (*a*) that the

---

[3] R. B. Stein, personal communication, 1963.

central effect of the stimulus serving as decision axis approximates a linear function of stimulus intensity over the threshold range; (*b*) and that the neural noise arises not from independent, but from positively correlated sources. With these additions the model explains both the approximation to Weber's law which is obtained at high stimulus intensities, and the departures from the law which occur at low intensities, for visual intensity discrimination and other dimensions, as well as features of visual spatial and temporal summation.

REFERENCES

AGUILAR, M., & STILES, W. S. Saturation of the rod mechanism of the retina at high levels of stimulation. *Optica Acta*, 1954, **1**, 59–65.

BARLOW, H. B. Summation and inhibition in the frog's retina. *J. Physiol.*, 1953, **119**, 69–88.

BARLOW, H. B. Retinal noise and absolute threshold. *J. Opt. Soc. Amer.*, 1956, **46**, 634–639.

BARLOW, H. B. Increment thresholds at low intensities considered as signal/noise discriminations. *J. Physiol.*, 1957, **136**, 469–488. (a)

BARLOW, H. B. Noise and the visual threshold. *Nature*, 1957, **180**, 1405. (b)

BARLOW, H. B. Intrinsic noise of cones. In, *Visual problems of colour*. Vol. 2. National Physical Laboratory Symposium, No. 8. London: Her Majesty's Stationery Office, 1958. Pp. 617–630. (a)

BARLOW, H. B. Temporal and spatial summation in human vision at different background intensities. *J. Physiol.*, 1958, **141**, 337–350. (b)

BARLOW, H. B., FITZHUGH, R., & KUFFLER, S. W. Change of organization in the receptive fields of the cat's retina during dark adaptation. *J. Physiol.*, 1957, **137**, 338–354.

BAUMGARDT, E. Visual spatial and temporal summation. *Nature*, 1959, **184**, 1951–2.

BIRDSALL, T. G. The theory of signal detectability. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955.

BLACKWELL, H. R. Contrast thresholds of the human eye. *J. Opt. Soc. Amer.*, 1946, **36**, 624–643.

BOUMAN, M. A., & VAN DER VELDEN, H. A. The two-quanta explanation of the dependence of the threshold values and visual acuity on the visual angle and the time of observation. *J. Opt. Soc. Amer.*, 1947, **37**, 908–919.

BRINDLEY, G. S. *Physiology of the retina and the visual pathway*. London: Edward Arnold, 1960.

CANE, V., & GREGORY, R. L. Noise and the visual threshold. *Nature*, 1957, **180**, 1404–1405.

CATTELL, J. M. On errors of observation. *Amer. J. Psychol.*, 1893, **5**, 285–293.

CROZIER, W. J. On the variability of critical illumination for flicker fusion and intensity discrimination. *J. gen. Physiol.*, 1936, **19**, 503–522.

CROZIER, W. J., & HOLWAY, A. H. On the law for minimal discrimination of intensities. I. *Proc. Nat. Acad. Sci.*, 1937, **23**, 23–28.

DE VRIES, H. The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. *Physica*, 1943, **10**, 553–564.

DIERCKS, K. J., & JEFFRESS, L. A. Interaural phase and the absolute threshold for tone. *J. Acoust. Soc. Amer.*, 1962, **34**, 981–984.

FECHNER, G. T. *Elemente der Psychophysik*. Leipzig: Breitkopf & Hartel, 1860.

FITZHUGH, R. The statistical detection of threshold signals in the retina. *J. gen. Physiol.*, 1957, **40**, 925–948.

GRAHAM, C. H., BROWN, R. H., & MOTE, F. A. The relation of size of stimulus and intensity in the human eye: I. Intensity thresholds for white light. *J. exp. Psychol.*, 1939, **24**, 555–573.

GRANIT, R. *Receptors and sensory perception*. New Haven: Yale Univer. Press, 1955.

GREGORY, R. L. An experimental treatment of vision as an information source and noisy channel. In E. C. Cherry (Ed.), *Information theory. Third London symposium*. London: Butterworths, 1956. Pp. 287–298.

GUILFORD, J. P. A generalized psychophysical law. *Psychol. Rev.*, 1932, **39**, 73–85.

HARTLINE, H. K. The effects of spatial summation in the retina on the excitation of the fibers of the optic nerve. *Amer. J. Physiol.*, 1940, **130**, 700–711.

HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

Hirsh, I. J. The influence of interaural phase on interaural summation and inhibition. *J. Acoust. Soc. Amer.*, 1948, **20**, 536–544.

Holway, A. H. On the precision of photometric observations. *J. Opt. Soc. Amer.*, 1937, **27**, 120–123.

Holway, A. H., & Crozier, W. J. On the law for minimal discrimination of intensities. II. *Proc. Nat. Acad. Sci.*, 1937, **23**, 509–515.

Jeffress, L. A., Blodgett, H. C., Sandel, T. T., & Wood, C. L. Masking of tonal signals. *J. Acoust. Soc. Amer.*, 1956, **28**, 416–426.

Kuffler, S. W., Fitzhugh, R., & Barlow, H. B. Maintained activity in the cat's retina in light and darkness. *J. gen. Physiol.*, 1957, **40**, 683–702.

Le Grand, Y. *Light colour and vision.* London: Chapman & Hall, 1957.

Neyman, J., & Pearson, E. S. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. A*, 1933, **231**, 289–337.

Pirenne, M. H. Physiological mechanisms of vision and the quantum nature of light. *Biol. Rev.*, 1956, **31**, 194–241.

Pirenne, M. H. Liminal brightness increments. In H. Davson (Ed.), *The eye.* Vol. 2. *The visual process.* New York: Academic Press, 1962.

Rose, A. The sensitivity performance of the human eye on an absolute scale. *J. Opt. Soc. Amer.*, 1948, **38**, 196–208.

Rushton, W. A. H. The Rhodopsin density in the human rods. *J. Physiol.*, 1956, **134**, 30–46.

Steinhardt, J. Intensity discrimination in the human eye. I. The relation of $\Delta I/I$ to intensity. *J. gen. Physiol.*, 1936, **20**, 185–209.

Stone, M. Models for choice-reaction time. *Psychometrika*, 1960, **25**, 251–260.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

Tanner, W. P., & Swets, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

Thurstone, L. L. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368–389. (a)

Thurstone, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273–286. (b)

Treisman, M. The nature of the psychophysical law and its significance for the scaling of sensory magnitudes. *Acta psychol., Amsterdam*, 1961, **19**, 213–214.

Treisman, M. Psychological explanation: The "private data" hypothesis. *Brit. J. Phil. Sci.*, 1962, **13**, 130–43.

Treisman, M. Laws of sensory magnitude. *Nature*, 1963, **198**, 914–915. (a)

Treisman, M. Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychol. Monog.*, 1963, **77** (13, Whole No. 576). (b)

Treisman, M. The effect of one stimulus on the threshold for another: An application of signal detectability theory. *Brit. J. statist. Pscyhol.*, 1964, in press. (a)

Treisman, M. Sensory scaling and the psychophysical law. *Quart. J. exp. Psychol.*, 1964, **16**, 11–22. (b)

Treisman, M. Signal detection theory and Crozier's law: Derivations of a new sensory scaling procedure, 1964, in press. (c)

Upton, M., & Crozier, W. J. On auditory intensity discrimination. *Proc. Nat. Acad. Sci.*, 1936, **22**, 417–420.

van der Velden, H. A. Over het aantal lichtquanta dat nodig is voor een lichtprikkel bij het menselijk oog. *Physica*, 1944, **11**, 179–189.

Willmer, E. N. The relationship between area and threshold in the central fovea for lights of different colours. *Quart. J. exp. Psychol.*, 1950, **2**, 53–59.

Woodworth, R. S. Professor Cattell's psychophysical contributions. *Arch. Psychol.*, New York, 1914, **4**, No. 30, 60–74.

(Received July 5, 1963)

Psychological Review
1964, Vol. 71, No. 4, 331–334

THEORETICAL NOTE

HOW SINCERE IS THE DOGMATIST?

GORDON STANLEY AND JOHN MARTIN
University of Western Australia

Rokeach has proposed two hypotheses as alternatives to the acquiescence explanation of the double-agreement phenomenon. The first of these (B_1) states that the respondent when replying to the reversed scale "sees a good reason why he should not tell the truth. . . . he (thus) tells the truth to the original statement and he deliberately lies to the reversed statement." This is interpreted as implying that responses to a reversed scale of dogmatism, authoritarianism, or anti-Semitism would correlate positively with scores on social desirability or lie scales. Empirical data are advanced to refute this hypothesis.

Rokeach (1963) has recently proposed two new hypotheses to explain acquiescence or the double-agreement phenomenon. Until now this phenomenon has been interpreted largely in terms of response style, but Rokeach's hypotheses return to a consideration of content. The first of these, Hypothesis B_1, states that,

> A person may agree with a statement and with its opposite because in both instances he reads them, comprehends them and responds to their content. He tells the truth in one case because he sees no reason why he should not; he deliberately lies in the second case because he sees a good reason why he should not tell the truth [Rokeach, 1963, pp. 304–305].

He considers that the subject agrees with the positively worded statement because its "antidemocratic tone" is not apparent. Agreement with such an item will indicate authoritarianism, anti-Semitism, or dogmatism as the case may be. When presented with the reversed item, the subject wishes to disagree with it but fears that so doing may produce a "bad" impression on the experimenter, consequently he lies in response to the reversed statement and produces double agreement.

Rokeach says that the converse of this hypothesis, viz. that the subject lies in response to the original items and tells the truth in response to the reversals is psychologically meaningless and hence improbable.

Hypothesis B_1 advances no explanation of the double-disagreement phenomenon (Peabody, 1961) but may be logically extended to say that the nonauthoritarian (pro-Semitic, nondogmatic) subject responds truthfully to the original and untruthfully to the reversal. If this extension is entertained it is apparent that a psychological explanation for these subjects' disagreement with the nonauthoritarian (pro-Semitic, etc.) item could also depend on their unwillingness to make a bad impression. It is not psychologically meaningless to say that subjects who are nonauthoritarian (etc.) may feel that these attitudes are not socially acceptable and may seek to conceal them by giving insincere replies to the negative item.

These hypotheses, together with other variations on the theme raised by Rokeach's Hypothesis B_1 may be conveniently tested by observing the correlation between responses to positive and reversed scales (of the specified type) and suitable measures of insincerity, e.g., social desirability and lie scales.

The various hypotheses are:

1. Insincerity appears only among the acquiescent group
   (a) in relation to the negative scale (Rokeach's Hypothesis B_1).
   (b) in relation to the positive scale (converse of above).

2. Insincerity appears only among the dissenting group
   (a) in relation to the negative scale.
   (b) in relation to the positive scale.
3. Insincerity appears among all subjects
   (a) in relation to the negative scale.
   (b) in relation to the positive scale.

The following data serve to test these hypotheses.

AN EMPIRICAL TEST

A reversed scale of dogmatism (D−) was constructed based on the same rationale as Rokeach (1956) and prima facie measuring the same psychological variable.[1] The wording of the original was departed from when it was considered that this allowed the production of a more satisfactory item (Christie, Havel, & Seidenberg, 1958), otherwise reversal was obtained by simple grammatical changes.

Research results (e.g., Bendig, 1962; Jackson & Messick, 1962) suggest that social desirability and lie scales are the best available measures of insincerity.

In this study the Martin Social Desirability (SDB) Scale, construction of which is being reported elsewhere, was used as the indicator of social desirability. This scale employs 23 positive and 23 negative items balanced for acquiescence. Martin has found reliabilities (Cronbach's alpha) for this scale of .84, .77, .75, and .70 with different groups. The lie (L) scale of the MPI (Gibson, 1962) was used as an indicator of faking or lying. Gibson reported the split-half reliability of this scale to be .66 (Guttman's formula). Reliabilities found for the present sample are given in Table 1.

These three measures, together with Rokeach's Dogmatism scale (D+) were administered to Psychology I students at the University of Western Australia ($N = 127$). Tests were administered in a special test program, the order of administration of D+ and D− being counterbalanced and separated in time by approximately 1 hour. D− was scored in reverse so that scores represent dogmatism.

TABLE 1

SHOWING CORRELATIONS BETWEEN POSITIVE AND REVERSED SCALES OF DOGMATISM, BALANCED (+ve AND −ve) DOGMATISM, SOCIAL DESIRABILITY, AND LIE SCALE

($N = 127$)

	SDB	D+	D−	DB^1	L
SDB	(.69)	−.234**	−.035	−.178*	+.397**
D+		(.74)	.174*	.801**	.148
D−			(.60)	.674**	−.136
DB^a					.005
L					(.63)

Note.—Figures in brackets are reliabilities (Cronbach's alpha).

[a] DB = balanced dogmatism scale.

* $p \leq .05$.

** $p \leq .01$.

RESULTS

When the data are analyzed for the group as a whole (Table 1) it is found that: (*a*) the expected inconsistency of response (whether due to acquiescence or dissent) appears when D+ and D− are compared. This is attested by the low positive correlation between these scales. It accords with the results found by Peabody (1961) using an independently constructed reversed scale of dogmatism. (*b*) D− has no relationship with SDB, and its relationship with the L scale, although in the direction predicted by Rokeach, is not significant. (*c*) the relationship of D+ with SDB is negative and highly significant, so that the subject answering in the socially desirable direction tends to disagree with the positive dogmatism item. The relationship of D+ with the L scale is positive but not significant.

The subjects were then divided at the means of D+ and D− to form four groups, two being characterized by consistent-to-content responses (but of high and low value) and two by predominantly stylistic responses (acquiescers and dissenters). Correlations between the relevant measures are shown separately for all four groups (Table 2). Only one of these coefficients is significant; that between D+ and SDB in the low-dogmatic

[1] Copies of the scale are available from the authors on request.

TABLE 2

SHOWING CORRELATIONS WITHIN HIGH-ACQUIESCENT, LOW-ACQUIESCENT, HIGH-DOGMATIC, AND LOW-DOGMATIC GROUPS

	D(+)	D(−)
High-acquiescent group ($N = 34$)		
SDB	−178	−082
Lie scale	.152	−014
Low-acquiescent (dissenters) group ($N = 27$)		
SDB	−291	−052
Lie scale	−359	−239
High-dogmatic group ($N = 32$)		
SDB	−308	.120
Lie scale	.153	−130
Low-dogmatic group ($N = 34$)		
SDB	−339*	−091
Lie scale	.014	.189

* $p \leq .05$.

group, and this is in the negative direction. Among the dissenting group the correlation between D+ and Lie scale approaches significance. This is also in the negative direction. Among the acquiescent group, which is the specified group for Rokeach's hypothesis, the correlations between D— and the measures of insincerity are negligibly different from zero.

CONCLUSIONS

Clearly if any hypothesis is supported by these results it is Hypothesis 3(b). Insincerity is related to low scores on the positive Dogmatism scale, the relationship being most pronounced among the low-dogmatic group and least among the acquiescent group. In view of the possibility that social desirability scales measure only the most popular form of faking (Martin & Stanley, 1963), this conclusion may require qualification.

There is no evidence to support Rokeach's Hypothesis B_1.

REVERSED SCALE OF DOGMATISM

1. The United States and Russia have many things in common.
2. You cannot have a democracy in which one section of the community has most of the power.
3. It is never necessary to restrict what people may say in public discussion of politics.
4. When a man believes something he must make himself familiar with all the arguments against it.
5. Man has within himself the power to control his destiny.
6. The world is full of warm human relationships.
7. In general most people show consideration for others.
8. I would resent assistance with my personal problems.
9. Optimism about the future is man's natural state.
10. What isn't done today can equally well be done tomorrow.
11. I never find it necessary to withdraw from a discussion no matter how heated.
12. In discussions people generally don't have much difficulty in understanding what I say.
13. Even in a heated discussion I always pay careful attention to the arguments of others.
14. It is not worth sacrificing your life to become a hero.
15. I have never wished to be famous.
16. Doing important work is not the main thing in life.
17. Even if I were given all the breaks, I would not be able to do anything of great benefit to the world.
18. A great proportion of the best thinkers have not had their achievements recorded in history.
19. I never take a dislike to people because of what they stand for.
20. It's possible to really live without believing in any great cause.
21. Life can be meaningful without devotion to ideals or causes.
22. All of the philosophies which exist in this world have some truth in them and probably not one is totally correct.
23. A person who is enthusiastic about many things is likely to develop greater character.
24. It is always best to meet our political opponents half-way.

25. Differences of opinion in religion are more apparent than real.
26. In these present days everyone should look to their own happiness.
27. A person who attacks publicly the people who believe in the same things he does shows personal integrity.
28. It is never necessary to be on guard against ideas no matter where they may originate.
29. Many differences of opinion can sustain a group.
30. Truth is so elusive that no one can say when he has it.
31. I like people to stick to their guns, even when they are wrong.
32. I think none the worse of a person for being concerned chiefly with his own pleasure.
33. In order to get published nowadays an idea has to be good.
34. To know the truth about what is going on we cannot rely simply on experts or leaders.
35. One should form one's own opinions, even if they conflict with those which others regard more highly.
36. The best way to live is to make friends and associates whose tastes and beliefs are different to one's own.
37. It is the present that counts—the future will look after itself.
38. If a man is to accomplish his mission in life he should never take chances.
39. People are very much aware of the shortcomings in our way of life.
40. Most people show a lot of foresight when it comes to something which affects them personally.

REFERENCES

BENDIG, A. W. A factor analysis of social desirability, defensiveness, lie and acquiescence scales. *J. gen. Psychol.*, 1962, **66**, 129-136.

CHRISTIE, R., HAVEL, J., & SEIDENBERG, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 142–159.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrica,* 1962, **16**, 297–334.

GIBSON, H. B. The lie scale of the Maudsley Personality Inventory. *Acta psychol.*, 1962, **20**, 18–23.

JACKSON, D. N., & MESSICK, S. Response styles and the assessment of psychopathology. In S. Messick & J. Ross (Eds.), *Measurement in personality and cognition.* New York: Wiley, 1962. Ch. 8.

MARTIN, J., & STANLEY, G. Social desirability and the Maudsley Personality Inventory. *Acta psychol.*, 1963, **21**, 260–264.

PEABODY, D. Attitude content and agreement set in scales of authoritarianism, dogmatism, anti-Semitism and economic conservatism. *J. abnorm. soc. Psychol.*, 1961, **63**, 1-11.

ROKEACH, M. The double agreement phenomenon: Three hypotheses. *Psychol. Rev.*, 1963, **70**, 304–309.

(Received November 4, 1963)

VOL. 71, NO. 5 SEPTEMBER 1964

PSYCHOLOGICAL REVIEW

CRITICAL PERIODS, STIMULUS INPUT, AND EMOTIONAL REACTIVITY:

A THEORY OF INFANTILE STIMULATION[1]

VICTOR H. DENENBERG[2]

Purdue University

Experiments, using rats and mice, do not support the critical period hypothesis that there are certain limited time periods in infancy during which a particular class of stimuli will have profound effects upon subsequent behavior. Where findings are consistent with the hypothesis, further research has shown that the "critical period" is a complex function of amount of infantile stimulation. The central hypothesis of this paper is that amount of stimulus input in infancy acts to reduce emotional reactivity in a monotonic fashion. From this it follows that an inverted U function should be obtained between amount of infantile stimulation and adult performance for tasks involving some form of noxious element and which are of "moderate" difficulty. For tasks which are "easy" or "difficult," the relationships between performance in adulthood and infantile stimulation should be monotonic, though opposite in slope. Data supporting this theory are discussed.

This paper has two purposes. The first is to question certain aspects of the critical period hypothesis (Scott, 1958, 1962). The second, and more important, purpose is to develop a somewhat different hypothesis: that stimulation in infancy reduces the organism's emotional reactivity; this reduction is a monotonic function of amount of stimulus input. In turn, this hypothesized change in emotional reactivity offers a useful mechanism by which one can explain a number of disparate and seemingly unrelated changes in adult performance.

The data to be considered, and the generalizations therefrom, come specifically from research with rats and mice which have received various forms of stimulation between birth and weaning. For more general reviews of early experience research and the critical period hypothesis see Denenberg (1962b) and Scott (1962).

CRITICAL PERIODS

Recently Scott (1962) has suggested that there are three major kinds of critical period phenomena: one for the formation of basic social relationships, a second concerned with optimal periods for learning, and a third involving infantile stimulation. This discussion is concerned only with the last of these.

---

[1] Much of the research described in this paper was supported, in part, by Grant M-1753 from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service, and by several grants from the Purdue Research Foundation.

[2] It is a pleasure to acknowledge the contributions of Mark W. Stephens, whose incisive comments, both conceptual and editorial, have significantly improved this paper.

There are at least two different ways in which the critical period hypothesis has been interpreted. One is that the same physical stimulation at different ages has different effects upon *S*; this says, simply, that *S*'s age is an important parameter and can scarcely be questioned. The second interpretation is that there are certain *limited* time periods in development during which a particular class of stimuli will have particularly profound effects and that the same stimulation before or after this interval will have little, if any, effect upon the developing organism. This approach stems directly from the embryological meaning of critical periods, and it is in this context that Scott (1958, 1962; Scott & Marston, 1950; Williams & Scott, 1953) has developed and described the critical period hypothesis. The term, critical periods, when used in this paper, will refer to this second interpretation.

Operationally, a test of the critical period hypothesis requires several experimental groups which receive the same stimulation at different ages and a control group which does not receive the stimulation at all. If some of the experimental groups do not differ from the control while others do, this may be construed as evidence supporting the critical period hypothesis. (On the other hand, any significant differences among the experimental groups are support for the first meaning of the hypothesis.)

Although it is obvious that the age at which *S* is stimulated is an important parameter, when one carefully examines experiments investigating critical periods in infancy in rats and mice, there is little evidence supporting the hypothesis that stimulation must occur only during certain delimited time intervals to affect the organism's subsequent behavior. Interestingly, the only unequivocal demonstrations of critical periods are to be found in papers which used physiological, rather than behavioral, endpoints (Bell, Reisner, & Linn, 1961; Levine & Lewis, 1959a).

The general conclusion noted above might not be apparent from a casual inspection of the literature. There are several studies which, if taken in isolation, appear to support the hypothesis. For example, two early experiments on critical periods (Denenberg, 1958, 1960) seemed to support the concept. In the first study mice initially conditioned between 20 and 40 days of age had essentially the same reconditioning scores at 50 days, and all these scores were significantly higher than either a 50-day control group or a group which was initially conditioned at 16 days. These data suggested a critical period around 20 days of age such that conditioning experience prior to that date did not affect subsequent reconditioning performance; experience after that date appeared to have the same effect upon reconditioning, regardless of the original age of conditioning or the actual amount of conditioning. A subsequent study with older *S*s (Denenberg, 1960) established that it was indeed the age of the animals rather than the time interval between conditioning and reconditioning which mediated the effects.[3]

---

[3] A number of investigators have also shown, with the rat, that stimulation in infancy has different consequences as compared to equivalent stimulation in adulthood. (These include handling in infancy and early adulthood, Levine, 1956; raising rats in groups of 6 or 12 during infancy or adulthood, Seitz, 1954; subjecting rats to handling and auditory stimulation in infancy and

However, another study with the mouse, though not directly comparable to the two above, disclosed some complexities concerning critical period phenomena which cast doubt on the degree of generality of the hypothesis. Denenberg and Bell (1960) studied avoidance learning in adulthood as a joint function of age of infantile stimulation (i.e., critical periods), intensity of stimulation in infancy, and intensity of electric shock during adult performance. These latter two parameters had been held constant in prior research. Analyses of the learning data revealed that critical periods were not significant as a main effect; but the Periods × Adult Shock and the Periods × Adult Shock × Infantile Stimulation interactions were significant. The disturbing thing was that the age variable (i.e., critical periods) was dependent upon the particular level of stimulation used in the adult learning task. By the selection of certain combinations of infantile stimulation levels and adult shock levels one could "prove" that stimulation during certain "critical periods" could facilitate, have no effect, or interfere with avoidance learning in adulthood. The phenomena, then, appear to be far more complex than previously recognized in the generalizations afforded by the critical period hypothesis.

These data suggest that any single "critical period" is, at least in part, dependent upon the parameter of stimulus intensity. Thus, Denenberg's prior studies (1958, 1960) should be looked upon with caution since different shock levels may have yielded very different findings.

Critical period research with the rat has generally used the procedure called *handling* rather than shock. This consists of removing the pups from the home cage, and placing the young into containers (e.g., cans containing wood shavings, small wooden compartments, grid-floored boxes). The pups are left there for a short duration (generally 2 to 3 minutes though some *E*s have used intervals of 8 minutes or longer) and are then returned to the home cage. This is generally done once a day for different prescribed periods.

In a series of experiments Denenberg and Karas (1960, 1961) found that rats handled for the first 10 days of life were superior with respect to body weight, avoidance learning, and survival capability to ones handled during the second 10 days. In general, both 10-day groups were superior to *S*s which had been handled for 20 days in infancy.

These data suggested that it might be possible to handle *S*s for fewer than 10 days and possibly isolate specific critical periods related to different dependent variables. Therefore Denenberg (1962a) handled different groups of rats for 3- or 5-day intervals at different ages during the first 10 days of life. Measures of body weight, avoidance learning behavior, and survival time, with one exception, did *not* support the critical period hypothesis. The exception was that *S*s handled for the first 5 days of life were not significantly better than controls in avoidance learning while *S*s handled on Days 6–10 were superior to controls and equal to *S*s handled on Days 1–10.

Though the avoidance learning findings are compatible with the critical period hypothesis, another possibility, based upon the Denenberg and Bell mouse study, is that the failure to modify avoidance learning by handling during the first 5 days was because the stimulation induced by handling was

---

adulthood, Spence & Maher, 1962a; and temporarily blinding or deafening rats in infancy or later life, Wolf, 1943.)

not sufficiently intense to affect *S*. Denenberg and Kline (1964) tested this hypothesis by giving different groups of rats electric shock (0.2 milliampere) on Days 1–5, 1–3, 3–5, 2, or 4. The shocked groups (with the exception of the group shocked only on Day 2) were significantly better in avoidance learning than a nondisturbed control group.

In conclusion, experimental research with the mouse and rat investigating stimulation between birth and weaning has established that the critical period hypothesis is not sufficient to account for the findings. In instances where the data are consistent with the hypothesis further study has found that the "critical period" is a complex function of the parameter of stimulus intensity. Research to date indicates that, for the rat and mouse at least, there may be as many "critical periods" as there are combinations of independent variable parameters and dependent variable measures (Denenberg, 1962a). Meyers (1962) and Lindholm (1962) have recently arrived at a similar conclusion.

Then should the term "critical periods" be abandoned when discussing the effects of infantile stimulation? Not at the present time, certainly. The concept has had, and will continue to have, great heuristic value. However, conclusions concerning critical periods must be limited to the particular experimental operations involved until enough empirical evidence has been obtained to indicate the limits of generalization. In terms of current research strategy it appears more reasonable to study the functional relationships among various classes of independent and dependent variables between birth and weaning than to design experiments which try to isolate critical periods during this crucial stage of development.

THEORY OF INFANTILE STIMULATION AND ADULT PERFORMANCE

One approach to the problem of understanding the effects of stimulation in infancy is by means of the critical period hypothesis. A different approach is to examine the diverse research findings to see if any common factor can be discerned. Such an examination of the literature leads to the conclusion that stimulation administered between birth and weaning brings about a reduction in "emotional reactivity." Furthermore, the greater the stimulation in infancy, the less emotional *S* will be in adulthood. The central hypothesis of this paper is that *emotional reactivity is reduced as a monotonic function of amount of stimulus input in infancy.*

Before discussing this hypothesis in detail, and showing some of its consequences, it is first necessary to define "emotional reactivity" and show that this intervening variable is modified by stimulation in infancy. This will be done by showing, on the stimulus side, that different methods of stimulating *S* in infancy have similar consequences in adulthood; and, on the response side, by showing that different operational measures of emotional reactivity yield internally consistent results (Miller, 1959). The two most common methods of stimulating *S*s in infancy have been handling and shocking. Emotional reactivity has been measured by procedures as divergent as open-field testing, consummatory behavior, behavior in a learning situation, and behavior when given an opportunity to emerge from the home cage.

Handling and Emotional Reactivity

Open-field Behavior. In two experiments Denenberg and Morton (1962a) found that rats handled daily between birth and weaning were significantly more active and had a significantly smaller defecation rate than nonhandled controls. Similar findings were obtained by Denenberg and Whimbey (1963). Rats handled for the first 10 or 20 days of life were found to be significantly more active and to defecate significantly less than nonhandled controls (Denenberg, Morton, Kline, & Grota, 1962). In a study relating infantile stimulation to age of testing Denenberg and Smith (1963) handled or shocked rats during the second 10 days of life and tested independent groups at 50, 100, 150, or 200 days of age. Both groups which received stimulation in infancy were significantly more active than the controls and had a significantly lower defecation percentage. The open-field test was also used by Schaefer (1963), but his index of emotionality was the amount of time spent crouching during the 2 minutes after presentation of a sharp click. Handled *S*s crouched significantly less than nonhandled controls.

Consummatory Behavior. Levine (1957, 1958) handled, shocked, or did not disturb rats between birth and weaning. When adult, the *S*s were deprived of water for 18 hours and then given the opportunity to drink. Levine (1957),

> hypothesized that since deprivation constitutes a novel internal stimulus complex for nonhandled *S*s, the novelty should result in greater emotional disturbance and produce reduced water consumption following a period of deprivation [p. 609].

In both experiments handled and shocked *S*s consumed significantly more water than nondisturbed controls. In a similar type of experiment Spence and Maher (1962a) also found that handled *S*s consumed significantly more water than controls.

Emergence-from-Cage Behavior. Hunt and Otis (1963) exposed their experimental rats to a variety of stimulus conditions in infancy, including handling. In their first two experiments their measure of emotionality was a "timidity" test. The home cage of an *S* which had been deprived of food and water for 22 hours was placed on an open alley upon which were placed pellets of food. Handled *S*s were found to emerge significantly further into the runway than nondisturbed controls. In their third experiment they recorded *S*'s behavior when *E* opened the cage door to attach a plastic milk cup to it, scoring *S* with respect to the degree of "boldness" manifested. Subjects not disturbed in infancy were found to be significantly more timid than those receiving stimulation in infancy.

Behavior in a Learning Situation. Several indexes of emotional behavior were recorded by Levine (1956) while testing handled and nonhandled *S*s for avoidance learning. These were defecation and activity during habituation trials, percentage of *S*s freezing after the first shock trial, and the number of trials on which *S*s showed freezing behavior. On all four measures nonhandled controls were found to be significantly more emotional than *S*s handled in infancy.

Electric Shock and Emotional Reactivity

Open-field Behavior. Subjects receiving 0.25 milliampere of current for 3 minutes daily on Days 11–20 were significantly more active and defecated significantly less than nondisturbed controls when tested at 50, 100, 150, or 200 days of age (Denenberg & Smith, 1963). Rats given 3 minutes of either 0.5 milliampere or 0.8 milliampere shock on Day 4 of life were significantly more active than *S*s receiving the same stimulation on Day 2; the latter group did not differ from nondisturbed controls (Kline & Denenberg, unpublished). The effect of electric shock is measurable at weaning as well. Rats given 30 seconds of 0.2 milliampere current on Days 1 and 2 of life were significantly less emotional than nondisturbed controls at the time of weaning as measured by time to emerge into the open field and activity in the field (Denenberg, Carlson, & Stephens, 1962).

Consummatory Behavior. Rat pups were given 3 minutes of electrical stimulation (variable current from 0.10 milliampere to 0.37 milliampere, disregarding *S*'s resistance) from Day 1 through Day 20. These *S*s consumed significantly more water in adulthood than nondisturbed controls; the findings were replicated in a second experiment (Levine, 1957, 1958). A similar finding was obtained by Lindholm (1962) who shocked rats for the first 10 days, the second 10 days, the first 20 days of life, and did not disturb a control group. No specified shock level was used; instead, the shock was set high enough to make the *S*s move and squeal in the apparatus. The intensity of the shock was increased as *S* matured. When adult, the *S*s were deprived of water for 24 hours and then given an opportunity to drink. The *S*s shocked in infancy took significantly less time to initiate drinking and consumed significantly more water than the nondisturbed group. A subsequent analysis determined that the differences in consummatory behavior could all be attributed to the latency to approach the water tubes.

Stimulus Input and Emotional Reactivity: The Monotonicity Hypothesis

The data summarized above are sufficient to justify the conclusion that stimulation in infancy will result in a reduction in emotional reactivity in adulthood. It remains to be demonstrated that a monotonic relationship obtains between amount of stimulus input in infancy and emotional reactivity. Two experiments have explicitly tested this hypothesis (Denenberg, Carlson, & Stephens, 1962; Denenberg & Smith, 1963). In addition, there are three other experiments in the literature which offer support for the hypothesis (Levine, 1957, 1958; Lindholm, 1962).

Amount of stimulus input was varied by Denenberg and his co-workers by handling rats for 0 (control), 10, or 20 days in infancy. In adult open-field testing the controls had the highest defecation rate while the group handled for 20 days had the lowest rate; all differences were significant. In the second study (Denenberg & Smith, 1963) one group of rats received 3 minutes of shock daily from Day 11 through Day 20, a second group was placed on the unelectrified grid (handled group), and a third group was not disturbed. Figure 1 shows the percentage of *S*s in each group defecating before and after adult avoidance learning training. The greater the stimulus input in infancy, the smaller (significantly) was the percentage of defecators in adulthood. Furthermore, the shocked *S*s were the only

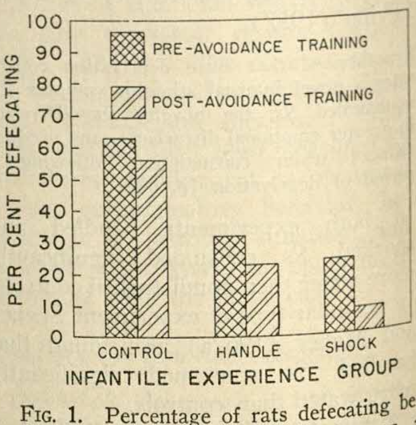


FIG. 1. Percentage of rats defecating before and after avoidance learning as a function of infantile stimulation (from Denenberg & Smith, 1963).

ones whose defecation percentage was significantly reduced following avoidance learning training.

Levine (1957, 1958) compared water consumption following 18 hours of thirst for *S*s which had been shocked, handled, or not disturbed in infancy. In both experiments the shocked *S*s consumed the most water, followed by the handled group, with the controls consuming the least. The difference between shocked and handled *S*s approached significance in the first study ($p < .10$) and was significant in the second study; nonhandled controls consumed significantly less than either experimental group in both studies.

In the Lindholm (1962) experiment rats were shocked for the first 10 days of life, the second 10 days, or the first 20 days, while a control group was not disturbed (his postweaning experimental group is not pertinent to this discussion). Consummatory behavior following 24 hours of thirst found that the controls had the greatest latency and consumed the least amount of water, followed by the two 10-day groups, with the 20-day group having the shortest latency and ingesting the greatest amount of water.

To summarize: Stimulus input in infancy has been varied by number of days of handling (0, 10, 20) (Denenberg et al., 1962), number of days of shock (0, 10, 20) (Lindholm, 1962), and form of stimulation (nondisturbed controls, handled, shocked) (Denenberg & Smith, 1963; Levine, 1957, 1958). Emotional reactivity has been measured by open-field behavior and consummatory behavior. In each instance the greater the amount of stimulus input in infancy the less was the level of *S*'s emotional reactivity in adulthood. Figure 2 presents an idealized curve showing this theoretical relationship. The negatively accelerated form of the curve is suggested by the findings of the five experiments cited above. In each instance the greatest change in behavior occurred between the control group and the group receiving the intermediate amount of stimulation.

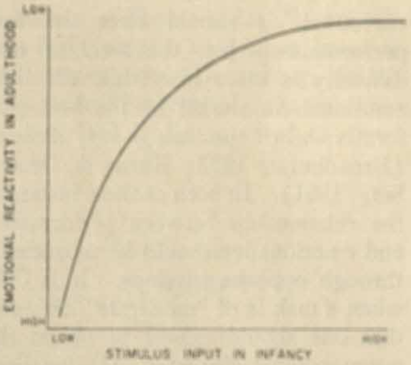


FIG. 2. Theoretical curve relating stimulus input in infancy to emotionality in adulthood.

Relationship between Emotional Reactivity and Adult Performance

When one examines the dependent variables used in studies of infantile stimulation, it is apparent that many of them contain some form of noxious element (avoidance learning, underwater swimming, thirst, starvation). Obviously, then, *S*'s emotionality or level of arousal will have a significant effect upon his performance in such tasks. It is reasonable to expect that there will be an optimal level of emotionality for efficient performance. As one moves away from this optimal level, performance should drop off, thus resulting in an inverted U function (cf. Hebb, 1955).

However, another parameter is needed. The Yerkes-Dodson Law (Broadhurst, 1957) posits that the optimal level of motivation for a task decreases as task difficulty increases. Assuming that the more emotional *S* is more motivated, it follows that highly emotional *S*s should have the best performance when the level of task difficulty is quite low while the least emotional *S*s should be the best performers when the task is very difficult (Broadhurst, 1957; Karas & Denenberg, 1961). In both of these instances the relationship between performance and emotionality should be monotonic, through opposite in slope. It is only when a task is of "moderate" difficulty that one should expect to obtain the nonmonotonic inverted U function. Figure 3 presents the theoretical relationship between adult performance and *S*'s emotionality for tasks of different levels of difficulty.[4]

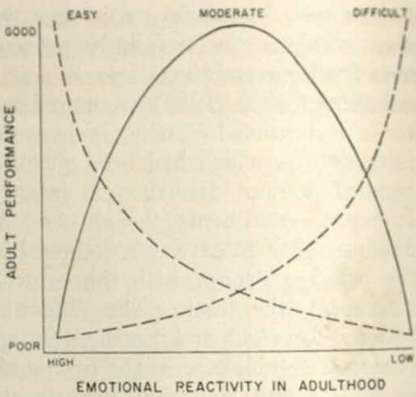


FIG. 3. Theoretical relationship between performance in adulthood and emotionality level for tasks of varying degrees of difficulty.

Some Relationships between Stimulus Input in Infancy and Adult Performance

A number of experiments have been carried out, the results of which are consistent with the predictions made by the theory depicted in Figures 2 and 3. For example, Karas and Denenberg (1961) explicitly tested the theory by assuming that the greater the number of days of handling in infancy, the greater the reduction in emotional reactivity in adulthood. In addition, they assumed that "spaced" handling would result in greater emotional reduction than "massed" handling. Experimental rats were space or mass handled for 10 or 20 days while controls were not disturbed. The criterion task was an underwater discrimination Y maze similar to the one used by Broadhurst (1957) who had shown that maze performance was related to *S*'s level of emotionality. The discrimination task was an "easy" one (86.6% correct choices, highly comparable to Broadhurst's value of 86.4% errorless trials on his "easy" task), thus leading to the deduction that the most emotional *S*s should exhibit the best performance with a monotonic decline in performance as emotional reactivity decreased (see Figure 3). Analysis of the swimming time scores found that the rank order of the five groups was as predicted, thus confirming the hypotheses.

Several studies have found an inverted U relationship between amount of stimulation in infancy and later performance. In the experiments by Denenberg (1962a) and Denenberg and Karas (1960, 1961) rat pups were handled for 0 (controls), 3, 5, 10, or 20 days in infancy. One can average the data for the 3-day, 5-day, and 10-day groups (irrespective of age of stimulation), add in the findings of the control and 20-day groups, and plot the functional relationship between number of days of handling in infancy and later performance. The averaging procedure acts to partially balance out the age of stimulation. One can, in other words, disregard "critical periods" between birth and weaning and examine amount of stimulation within that period. Such an analysis is shown in Figure 4 for 21- and 69-day body weight, avoidance learning, and survival time (Denenberg, 1962a). The general function of these data may be described by an inverted U curve.

---

[4] Still another parameter which may be relevant is rate of development. Denenberg and Karas (1959), in comparing the effects of handling upon the rat and mouse, point out that the mouse is a more rapidly developing organism and they suggest the hypothesis that the more rapid an organism's development, the greater the effect of infantile experience.

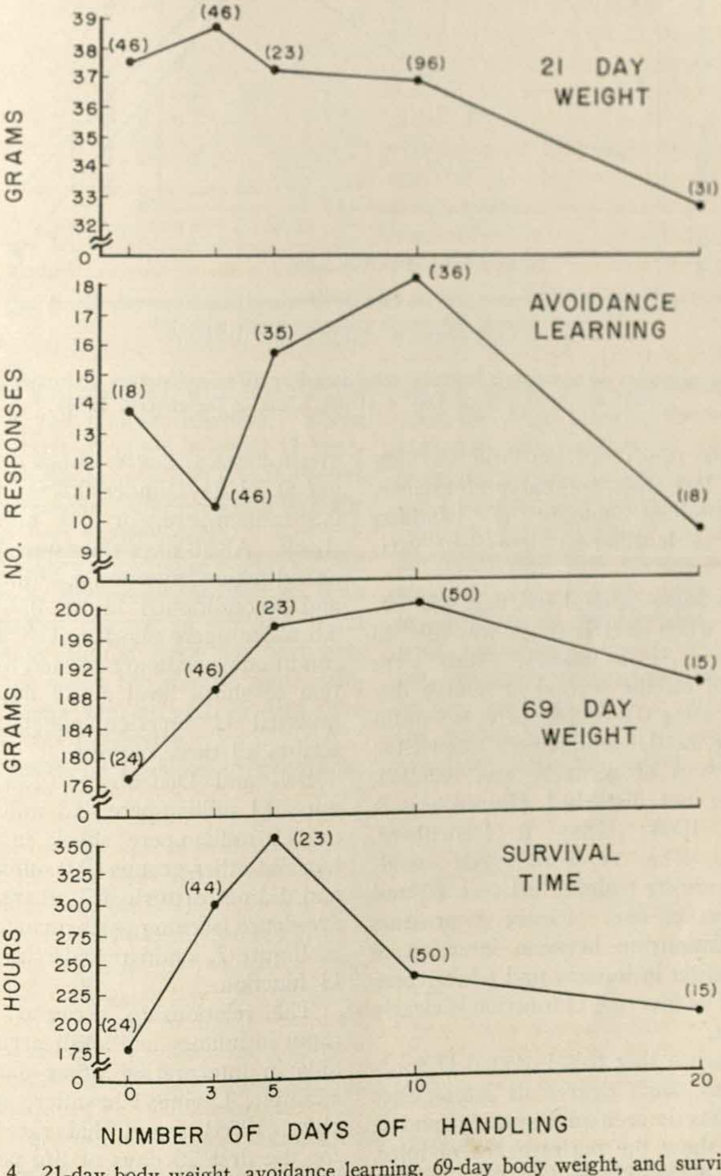


FIG. 4. 21-day body weight, avoidance learning, 69-day body weight, and survival time in rats as a function of number of days of handling in infancy. (Numbers in parentheses indicate the *N* per point, from Denenberg, 1962a.)

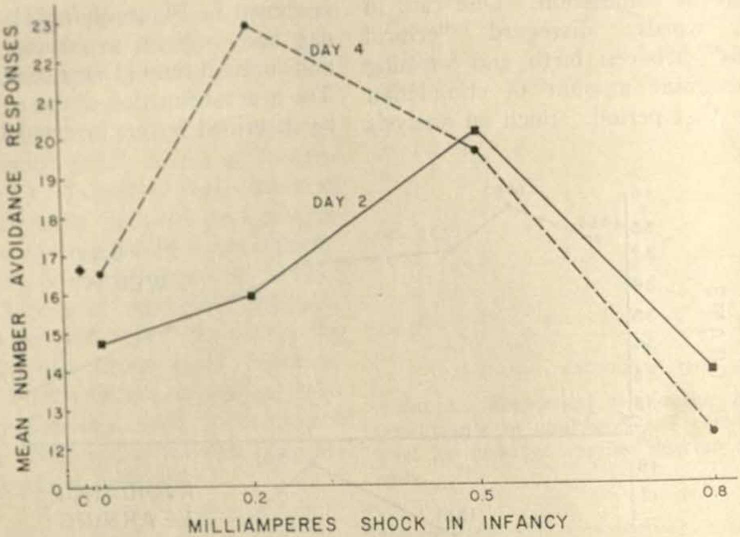


FIG. 5. Number of avoidance learning responses in rats as a function of infantile shock intensity on Day 2 or Day 4 (from Kline & Denenberg, 1964).

Both too much and too little handling led to less than optimal performance. The optimal amount of handling varied with different dependent variables.

This same general function was obtained when electric shock was used to stimulate *S*s in infancy. Rats were shocked on the second or fourth day of life using 0.2 milliampere, 0.5 milliampere, or 0.8 milliampere electricity; one group of controls was handled, another not disturbed (Denenberg & Kline, 1964; Kline & Denenberg, 1964). The *S*s were given avoidance learning training between 60 and 69 days of age. Figure 5 presents the relationship between intensity of stimulation in infancy and adult learning. The inverted U function is clearly present.

Evidence that this inverted U function has some degree of interspecies generality is seen in Figures 6 and 7, which show the performance of mice. Figure 6 summarizes an experiment by Denenberg (1959) in which *S*s were given classical buzzer-shock conditioning at 25 days under 0.2 milliampere, 0.5 milliampere, or 0.8 milliampere shock. At 50 days they were taken to an extinction criterion, split into thirds, and reconditioned under 0.2, 0.5, or 0.8 milliampere shock. The adult reconditioning data are plotted as a function of shock level at 25 days. The inverted U function adequately describes all three curves.

Bell and Denenberg (1963) gave mice 0.1 milliampere, 0.3 milliampere, or 0.5 milliampere shock in infancy, handled other groups (0.0 milliampere) and did not disturb still others. Adult avoidance learning performance, plotted in Figure 7, again reveals the inverted U function.

The relationship between stimulus input in infancy and adult performance aids in interpreting other data. For example, Levine, Chevalier, and Korchin (1956) found that rats handled for the first 20 days of life performed best in an adult avoidance learning task, followed by *S*s which had been

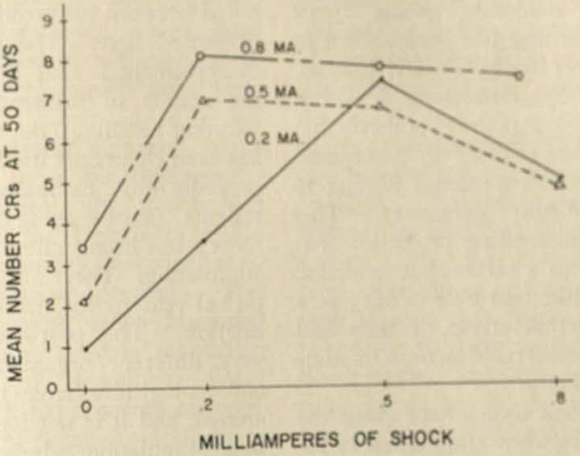


FIG. 6. 50-day conditioning scores in mice for three shock levels as a function of infantile shock intensity (from Denenberg, 1959).

handled and shocked, with nonhandled controls the poorest learners. These data exhibit a rough inverted U function with the group receiving the intermediate amount of stimulation exhibiting the best performance.

The theory depicted in Figures 2 and 3 leads to the prediction that *S*s with the greatest amount of stimulus input in infancy should have the best performance when the task is quite difficult. There are, as yet, no data relevant to this prediction.

It should be possible to generalize the parameter of "task difficulty" beyond the usual learning context. For example, the inverted U curves for body weight and survival time in Figure 4 were obtained after *S*s had undergone avoidance learning training. For *S*s which did *not* receive avoidance training, the controls lived longest, followed by *S*s handled for 10 days, with the 20-day group dying earliest. In other words, those data parallel the "easy" curve in Figure 3 while the results of the groups which had avoidance training follow along the

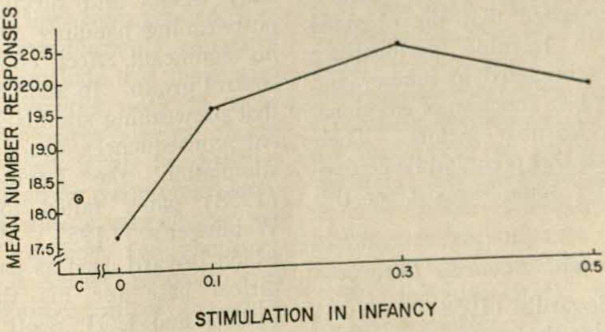


FIG. 7. Number of avoidance responses in mice as a function of stimulation in infancy. (C: nonstimulated controls; O: handled, nonshocked; 0.1, 0.3, 0.5: milliamperes of shock, from Bell & Denenberg, 1963.)

curve of the "moderate" group. Since avoidance learning has been shown to be a stressor (Brady, Porter, Conrad, & Mason, 1958; Denenberg & Karas, 1961), this suggests the hypothesis that stress experience, prior to a terminal stress, may act in a manner similar to the task difficulty parameter. This leads to the interesting prediction that *S*s subjected to a severe, but nonlethal stress, would be best able to survive a subsequent lethal stress, if they had received a considerable amount of stimulation in infancy.

It is pertinent to ask here about the relationship between stimulation in infancy and performance on a task which minimizes the emotional component. One would expect little or no relationship between infantile stimulation and later behavior. This is exactly what Denenberg and Morton (1962b) found in a series of three experiments. Handling in infancy did not have any effect upon problem-solving behavior as measured by the Hebb-Williams maze test. The authors concluded that,

> These findings suggest that preweaning stimulation such as shock and handling affects emotional processes but does not have any direct effect upon perceptual or problem-solving behavior [p. 1098].

Schaefer (1963) has obtained similar results.

Similarly, Spence and Maher (1962b) suggested that the obtained differences in learning performance between *S*s stimulated in infancy and controls may be a function of emotional rather than learning factors. Their experiment on this point, while equivocal, does offer some support for this position.

Comparisons with Bovard's Theory

Though Bovard's (1958) theory is concerned with the effects of handling upon viability, there are certain points of agreement and disagreement to be noted between his position and the one presented here. The major point of disagreement is that Bovard combines the results of preweaning and postweaning handling (the term "gentling" has been commonly used to describe the procedure of manually manipulating rodents postweaning), while this theory has been limited to preweaning stimulation. There are logical and empirical reasons for making such a distinction. The postweaning rat is a very different organism, behaviorally and biologically, from the preweaning animal, and it is not logical to assume that stimulation administered to the immature, recently born rat has the same consequences as the same stimulation administered to a weanling. The experiments cited in Footnote 3 offer general support for this conclusion. Clear evidence that handling and electric shock after weaning do not affect rats in the same manner as handling and shock before weaning can be found in the monograph by Brookshire, Littman, and Stewart (1961).

A paper directly relevant to this issue is the one by Levine and Otis (1958). They either handled or gentled rats before or after weaning; a control group was not disturbed. Preweaning handling or gentling had the same effect of significantly increasing body weight and survival rate while postweaning handling or gentling had no significant effect, compared to the control group. In addition to showing that preweaning stimulation has different consequences from postweaning stimulation, the Levine and Otis (1958) paper failed to substantiate Weininger's (1953, 1956) results which Bovard used as part of the empirical base for his theory. Ader (1959) and J. H. Scott (1955) have also failed to replicate Weininger's findings. Even the Hammett (1922) reference cited by Bovard is not supportive of his position, since the experimental data reported in that paper was not concerned with the effects of gentling rats. Hammett does refer back to his 1921 paper, and it is that portion of Hammett's paper which Bovard refers to in his reference to Hammett. Several other researchers have also cited the Hammett (1921) paper when discussing the positive effects of postweaning gentling. This is most interesting, since Hammett presents absolutely no data supporting such a conclusion. Hammett (1921) carried out five experiments concerning the effects of parathyroidectomy or thyroidectomy upon survival. The first three studies do not permit a valid evaluation of the effects of gentling because of lack of appropriate controls (though the experiments were valid for Hammett's purposes). The fourth and fifth experiments (see Hammett, 1921, p. 201) can be used to evaluate the effects of gentling since Hammett shifted the rats at weaning from one colony room to the other, thus balancing environmental conditions for the two different stocks of animals. The only significant effect in the fourth experiment was that the two colony stocks differed in genetic make-up; gentling failed to have any effect. And the fifth experiment contradicts this: all rats, regardless of genetic stock or whether or not they were gentled, survived the parathyroidectomy operation. One can only conclude that those who have cited Hammett's work as proof that gentling enhances survival ability have allowed their fertile imagination to inhibit their reading comprehension. All these experiments suggest that the postweaning handling (or gentling) phenomenon is rather tenuous with respect to modification of the rat's physiology.

There is some degree of agreement between Bovard's theory and the present one concerning preweaning stimulation. In so far as behavioral indices of emotional reactivity correlate with what Bovard (1958) calls stress, then the two theories are in agreement that "early [preweaning] handling raises the threshold for response to stress [p. 259]." Bovard, however, does not believe that handling is stressful while Levine (1956) and Denenberg (1959) have taken the opposite position. Bovard believes that "early stress lowers resistance to later stress [p. 260]." Since electric shock—a stimulus typically considered to be stressful—and handling have similar functional properties *vis à vis* emotional reactivity, the Levine and Denenberg position appears to be on firmer ground. Further evidence that stressful stimulation in infancy can have beneficial results, rather than the deleterious results predicted by Bovard, is discussed in the next section.

DISCUSSION

The hypothesis that emotionality decreases as a monotonic function of stimulus input may not appear reasonable at first glance because of our knowledge of conditioned fears, traumatic experiences in childhood, etc. It must be emphasized that this hypothesis is based upon, and is restricted to, data obtained on rodents which have been stimulated prior to weaning. Weaning has been used as a criterion point, not because of any clinical implications concerning mother-young separation, but because all of *S*'s senses are functioning by this time and because it is only after weaning that there is any evidence of long-term retention of a learned fear response (Campbell & Campbell, 1962; Denenberg, 1958, 1959; Lindholm, 1962). Furthermore, the consequences of handling or electric shock after weaning have been shown to be very different from their effects prior to weaning (Brookshire, Littman, & Stewart, 1961).

Scott (1962) has made a similar suggestion and has indicated that the transition point may occur at about 16 days of age, when the eyes open. Up to that age the critical factor mediating infantile stimulation is hypothesized by Scott to be the adrenal cortical stress mechanism while the psychological process of reduction of fear through familiarity is presumed to predominate between 17 days when the eyes open and 30 days.

The traditional assumption that traumatic experiences in infancy must inevitably have deleterious effects in later life is clearly not consonant with many of the studies cited above. Several other experiments, in fact, may also be noted which contradict this classical assumption. Baron, Brookshire, and Littman (1957) found that rats given 1.25 milliamperes of electric current continuously for 3 minutes on Days 20 and 21 were better escape and avoidance learners in adulthood than controls. Infant rats either placed on a laboratory shaker which oscillated 180 times per minute or shocked with 0.1 milliampere electricity were found to exhibit earlier maturation of the adrenal ascorbic acid depletion response to stress (Levine & Lewis, 1959b). Using the same dependent variable Schaefer, Weingarten, and Towne (1962) showed that placing rat pups into a refrigerator at 7° to 10° Centigrade also resulted in earlier maturation of the adrenal responses. Finally, Werboff and Havlena (1963) induced febrile convulsions in 3-day old rats by means of microwave diathermy. The *S*s so treated weighed significantly more at weaning and were significantly more resistant to audiogenic seizures.

Certain parallels may be noted between this theory and the arousal or activation theories of Hebb (1955), Duffy (1957), and Malmo (1959). All the theories are concerned with the intensity dimension of stimulation or arousal, and the inverted U function relating the intensity dimension to performance is common to all. The theory proposed here contributes to general arousal theory in two ways. First, intensity of infantile stimulation is specifically implicated as a major parameter affecting later differences in "chronic" or general level of arousal. Brookshire et al. (1961) have isolated what may be a similar phenomenon with the postweaning rat. They suggest that their "pure shock" residual factor of shock trauma may act to modify *S*'s arousal level. The second contribution is a methodological one: the techniques used in infantile stimulation are also procedures for experimentally generating individual differences in chronic arousal level. This opens up a new avenue of attack for those interested in individual differences in arousal level.

Though the proposed theory does account for a considerable number of experimental findings, contrary data are to be noted. For example, Spence and Maher (1962a) subjected rats to intense auditory stimulation between birth and weaning, handling a second group but did not give them auditory stimulation, moved the cages of a third group, and did not disturb a fourth group. Using water consumption as their index of emotionality no differences were found among the experimental groups though the stimulated *S*s consumed significantly more water than the undisturbed controls. No reason can be suggested to account for the failure to find a monotonic relationship between stimulus intensity in infancy and consummatory behavior.

Finally, this theory is not meant to be a substitute for the critical period hypothesis. In fact, the functional relationships are most clearly seen when *S*s are stimulated at different ages (but for the same number of days) and the date averaged so that age of stimulation is equated. Ultimately, any general theory of infantile stimulation will have to account for both age of stimulation and stimulus input and will have to relate these to the psychology and biochemistry of ontogeny (Levine, 1962).

REFERENCES

ADER, R. The effects of early experience on subsequent emotionality and resistance to stress. *Psychol. Monogr.*, 1959, **73**(2, Whole No. 472).

BARON, A., BROOKSHIRE, K. H., & LITTMAN, R. A. Effects of infantile and adult shock-trauma upon learning in the adult white rat. *J. comp. physiol. Psychol.*, 1957, **50**, 530–534.

BELL, R. W., & DENENBERG, V. H. The interrelationships of shock and critical periods in infancy as they affect adult learning and activity. *Anim. Behav.*, 1963, **11**, 21–27.

BELL, R. W., REISNER, G., & LINN, T. Recovery from electroconvulsive shock as a function of infantile stimulation. *Science*, 1961, **133**, 1428.

BOVARD, E. W. The effects of early handling on viability of the albino rat. *Psychol. Rev.*, 1958, **65**, 257–271.

BRADY, J. V., PORTER, R. W., CONRAD, D. G., & MASON, J. W. Avoidance behavior and the development of gastroduodenal ulcers. *J. exp. Anal. Behav.*, 1958, **1**, 69–73.

BROADHURST, P. L. Emotionality and the Yerkes-Dodson law. *J. exp. Psychol.*, 1957, **54**, 345–352.

BROOKSHIRE, K. H., LITTMAN, R. A., & STEWART, C. N. Residue of shock-trauma in the white rat: A three-factor theory. *Psychol. Monogr.*, 1961, **75**(10, Whole No. 514).

CAMPBELL, B. A., & CAMPBELL, E. H. Retention and extinction of learned fear in infant and adult rats. *J. comp. physiol. Psychol.*, 1962, **55**, 1–8.

DENENBERG, V. H. Effects of age and early experience upon conditioning in the C57BL/10 mouse. *J. Psychol.*, 1958, **46**, 211–226.

DENENBERG, V. H. The interactive effects of infantile and adult shock levels upon learning. *Psychol. Rep.*, 1959, **5**, 357–364.

DENENBERG, V. H. A test of the critical period hypothesis and a further study of the relationship between age and conditioning in the C57BL/10 mouse. *J. genet. Psychol.*, 1960, **97**, 379–384.

DENENBERG, V. H. An attempt to isolate critical periods of development in the rat. *J. comp. physiol. Psychol.*, 1962, **55**, 813–815. (a)

DENENBERG, V. H. The effects of early experience. In E. S. E. Hafez (Ed.), *The behaviour of domestic animals.* London: Bailliere, Tindall, & Cox, 1962. Pp. 109–138. (b)

DENENBERG, V. H., & BELL, R. W. Critical periods for the effects of infantile experience on adult learning. *Science*, 1960, **131**, 227–228.

DENENBERG, V. H., CARLSON, P. V., & STEPHENS, M. W. Effects of infantile shock upon emotionality at weaning. *J. comp. physiol. Psychol.*, 1962, **55**, 819–820.

DENENBERG, V. H., & KARAS, G. G. Effects of differential handling upon weight gain and mortality in the rat and mouse. *Science*, 1959, **130**, 629–630.

DENENBERG, V. H., & KARAS, G. G. Interactive effects of age and duration of infantile experience on adult learning. *Psychol. Rep.*, 1960, **7**, 313–322.

DENENBERG, V. H., & KARAS, G. G. Interactive effects of infantile and adult experiences upon weight gain and mortality in the rat. *J. comp. physiol. Psychol.*, 1961, **54**, 685–689.

DENENBERG, V. H., & KLINE, N. J. Stimulus intensity *vs.* critical periods: A test of two hypotheses concerning infantile stimulation. *Canad. J. Psychol.*, 1964, **18**, 1–5.

DENENBERG, V. H., & MORTON, J. R. C. Effects of environmental complexity and social groupings upon modification of emotional behavior. *J. comp. physiol. Psychol.*, 1962, **55**, 242–246. (a)

DENENBERG, V. H., & MORTON, J. R. C. Effects of preweaning and postweaning manipulations upon problem-solving behavior. *J. comp. physiol. Psychol.*, 1962, **55**, 1096–1098. (b)

DENENBERG, V. H., MORTON, J. R. C., KLINE, N. J., & GROTA, L. J. Effects of duration of infantile stimulation upon emotionality. *Canad. J. Psychol.*, 1962, **16**, 72–76.

DENENBERG, V. H., & SMITH, S. A. Effects of infantile stimulation and age upon behavior. *J. comp. physiol. Psychol.*, 1963, **56**, 307–312.

Denenberg, V. H., & Whimbey, A. E. Infantile stimulation and animal husbandry: A methodological study. *J. comp. physiol. Psychol.*, 1963, **56**, 877–878.

Duffy, E. The psychological significance of the concept of "arousal" or "activation." *Psychol. Bull.*, 1957, **64**, 265–275.

Hammett, F. S. Studies of the thyroid apparatus. I. The stability of the nervous system as a factor in the resistance of the albino rat to the loss of the parathyroid secretion. *Amer. J. Physiol.*, 1921, **56**, 196–204.

Hammett, F. S. Studies of the thyroid apparatus. V. The significance of the comparative mortality rates of parathyroidectomized wild Norway rats and excitable and non-excitable albino rats. *Endocrinology*, 1922, **6**, 221–229.

Hebb, D. C. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–254.

Hunt, H. F., & Otis, L. S. Early "experience" and its effects on later behavioral processes in rats: I. Initial experiments. *Trans. N. Y. Acad. Sci.*, 1963, **25**, 858–870.

Karas, G. G., & Denenberg, V. H. The effects of duration and distribution of infantile experience on adult learning. *J. comp. physiol. Psychol.*, 1961, **54**, 170–174.

Kline, N. J., & Denenberg, V. H. Qualitative and quantitative dimensions of infantile stimulation. Unpublished manuscript, Purdue University, Department of Psychology, 1964. (Ditto)

Levine, S. A further study of infantile handling and adult avoidance learning. *J. Pers.*, 1956, **25**, 70–80.

Levine, S. Infantile experience and consummatory behavior in adulthood. *J. comp. physiol. Psychol.*, 1957, **50**, 609–612.

Levine, S. Noxious stimulation in infant and adult rats and consummatory behavior. *J. comp. physiol. Psychol.*, 1958, **51**, 230–233.

Levine, S. The psychophysiological effects of early stimulation. In E. Bliss (Ed.), *Roots of behavior.* New York: Hoeber, 1962.

Levine, S., Chevalier, J. A., & Korchin, S. J. The effects of early shock and handling on later avoidance learning. *J. Pers.*, 1956, **24**, 475–493.

Levine, S., & Lewis, G. W. Critical periods and the effects of infantile experience on maturation of stress response. *Science*, 1959, **129**, 42–43. (a)

Levine, S., & Lewis, G. W. The relative importance of experimenter contact in an effect produced by extra-stimulation in infancy. *J. comp. physiol. Psychol.*, 1959, **52**, 368–369. (b)

Levine, S., & Otis, L. S. The effects of handling before and after weaning on the resistance of albino rats to later deprivation. *Canad. J. Psychol.*, 1958, **12**, 103–108.

Lindholm, B. W. Critical periods and the effects of early shock on later emotional behavior in the white rat. *J. comp. physiol. Psychol.*, 1962, **55**, 597–599.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

Meyers, W. J. Critical period for the facilitation of exploratory behavior by infantile experience. *J. comp. physiol. Psychol.*, 1962, **55**, 1099–1101.

Miller, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation and social learning. In S. Koch (Ed.), *Psychology, a study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

Schaefer, T., Jr. Early "experience" and its effects on later behavioral processes in rats: II. A critical factor in the early handling phenomenon. *Trans. N. Y. Acad. Sci.*, 1963, **25**, 871–889.

Schaefer, T., Weingarten, F. S., & Towne, J. C. Temperature change: The basic variable in the early handling phenomenon? *Science*, 1962, **135**, 41–42.

Scott, J. H. Some effects at maturity of gentling, ignoring, or shocking rats during infancy. *J. abnorm. soc. Psychol.*, 1955, **51**, 412–414.

Scott, J. P. Critical periods in the development of social behavior in puppies. *Psychosom. Med.*, 1958, **20**, 42–54.

Scott, J. P. Critical periods in behavioral development. *Science*, 1962, **138**, 949–958.

Scott, J. P., & Marston, M. V. Critical periods affecting the development of normal and maladjustive social behavior in puppies. *J. genet. Psychol.*, 1950, **77**, 25–60.

Seitz, P. F. D. The effects of infantile experiences upon adult behavior in animal subjects: I. Effects of litter size during infancy upon adult behavior in the rat. *Amer. J. Psychiat.*, 1954, **110**, 916–927.

Spence, J. T., & Maher, B. A. Handling and noxious stimulation of the albino rat: I. Effects on subsequent emotionality. *J. comp. physiol. Psychol.*, 1962, **55**, 247–251. (a)

SPENCE, J. T., & MAHER, B. A. Handling and noxious stimulation of the albino rat: II. Effects on subsequent performance in a learning situation. *J. comp. physiol. Psychol.*, 1962, **55**, 252–255. (b)

WEININGER, O. Mortality of albino rats under stress as a function of early handling. *Canad. J. Psychol.*, 1953, **7**, 111–114.

WEININGER, O. The effects of early experience on behavior and growth characteristics. *J. comp. physiol. Psychol.*, 1956, **49**, 1–9.

WERBOFF, J., & HAVLENA, J. Febrile convulsions in infant rats, and later behavior. *Science*, 1963, **142**, 684–685.

WILLIAMS, E., & SCOTT, J. P. The development of social behavior patterns in the mouse, in relation to natural periods. *Behaviour*, Leiden, 1953, **6**, 35–64.

WOLF, A. The dynamics of the selective inhibition of specific functions in neurosis: A preliminary report. *Psychosom. Med.*, 1943, **5**, 27–38.

(Received June 10, 1963)

Psychological Review
1964, Vol. 71, No. 5, 352–379

THE PROBLEM OF ATTENTION IN RESEARCH AND THEORY IN SCHIZOPHRENIA[1]

JULIAN SILVERMAN

Veterans Administration Hospital, Palo Alto

Studies of schizophrenia have been plagued, perhaps more so than any other area of psychological research, by the confounding effects of impaired attention on laboratory and clinical behaviors. This paper analyzes the perceptual response determinants of attentiveness and inattentiveness in schizophrenia. A "cognitive control" theory of attention is presented and elaborated through the use of research and clinical data. An attention-response mechanism termed "sensory input processing–ideational gating" is then formulated in an attempt to explain some heretofore incomprehensible aspects of schizophrenic disorders.

Psychiatrists have long regarded "impaired attention" as a predominant feature of schizophrenia (Kraepelin, 1919). Recently a number of psychological studies have again highlighted the significance of attention defects in schizophrenics. Weckowicz and Blewett (1959), in a study on size constancy and abstract thinking concluded:

> the abnormalities of thinking and perception in schizophrenic patients can be described as an inability "to attend selectively" or "to select relevant information" [p. 927].

Donahoe, Curtin, and Lipton (1961) after studying verbal learning stated:

> preliminary work strongly suggests that performance deteriorates inordinately if the schizophrenic is not allowed to program the stimuli himself [p. 557].

Shakow (1962), reviewing a number of studies by him and his colleagues on set impairment in schizophrenia, theorized:

> It is as if, in the scanning process which takes place before the response to a stimulus is made, the schizophrenic is unable to select out the material relevant for optimal response [p. 9].

While the above statements are far from startling, there is a paucity of researches which elaborate upon the above notions. Most of the psychological studies of schizophrenics report again and again and again that schizophrenics are sicker and do more poorly than normals on "experimental tasks" or conclude that some schizophrenics are not as sick as other schizophrenics. In the following pages, several hypotheses about the process of attention in schizophrenics will be presented. With these hypotheses it is possible to integrate and reinterpret a variety of disparate laboratory and clinical studies of schizophrenics and to provide a framework for future research on attention problems in schizophrenia.

Attention can be described in terms of both *extensive* and *selective* factors which seem to regulate the reception and utilization of stimulus input (e.g.,

---

[1] This paper was begun during the term of a predoctoral fellowship at the University of Michigan and was supported, in part, by Public Health Service Fellowship MF 17981 from the National Institute of Mental Health, Public Health Service. It was completed during the period of a Veterans Administration Psychology Research Associateship at the Palo Alto Veterans Administration Hospital.

The author wishes to express his appreciation to Enoch Callaway, Norman Garmezy, Paul McReynolds, and especially to Margaret Singer for their most helpful advice and criticisms.

Berlyne, 1960; Santos, Farrow, & Haines, 1963; Solley & Murphy, 1960). The extensive aspect of attention refers primarily to the degree to which stimuli are sampled from the environment. It is inferred from laboratory tasks which emphasize perceptual scanning (e.g., Gardner, Holzman, Klein, Linton, & Spence, 1959). Selectiveness of attention involves the *articulation* of a stimulus field into salient and irrelevant cues. It is inferred from tasks which necessitate attention to certain segments of a stimulus field and simultaneous inhibition of attention to other segments of the field. This paper accounts for striking individual differences between schizophrenics on laboratory measures of perceptual scanning and field articulation and demonstrates systematic relationships between attention-response styles and schizophrenic symptom patterns and pre-illness histories. Extreme forms of scanning and field articulation, minimal or excessive, are shown to characterize the attention-response dispositions of most schizophrenics.

The significance of these response dispositions, derive, on the one hand, from their relationships to personality structure and, on the other hand, through their use in the analysis of basic perceptual orientations to the environment. A detailed consideration of these interrelationships leads to the postulation of an "ideational-gating mechanism" in schizophrenics which functions to filter or to distort disturbing connotative environmental inputs.

Subgroups within Schizophrenia

The designation of diagnostic criteria which differentiate among the diverse forms of schizophrenia is sorely needed. Undesirable heterogeneity within current diagnostic classifications is grossly evident in many studies of schizophrenia in which both interindividual and intraindividual variability are discouragingly great. In the course of this review, an analysis of attention response styles will be used to delineate sources of correlation and also of confusion among the most prominent categorizations of schizophrenia. It will be shown that, despite reports of unreliability of diagnosis, there is much to recommend the use of classical Kraepelinian diagnosis of schizophrenic subtypes in selecting subjects for research. This will apply particularly to the paranoid-nonparanoid differentiation (e.g., Shakow, 1962). The utility of the process-reactive distinction also will be shown to be extremely important (e.g., Garmezy & Rodnick, 1959). Additionally, it will become clear that Kraepelinian subgroupings and the process-reactive dichotomy are intertwined with chronicity factors and physiological arousal patterns in complex ways but that the lack of adequate experimental controls in many studies has served to obscure the nature of such complex interactions.

Paranoid-Nonparanoid Distinction. Numerous experimental and clinical studies report significant differences between paranoid and nonparanoid schizophrenics on a variety of neurophysiological, perceptual, thought disorder, and personality variables (e.g., Foulds & Owen, 1963; Havener & Izard, 1962; Jackson, 1960, Kraepelin, 1919; Lynn, 1963; Raush, 1952; Sexton, 1945; Silverman, 1963a, 1964a, 1964b; Snyder, 1961; Voth, A. C., 1947). In addition, several recent papers have pointed to the theoretical and research confusion which occurs when paranoid schizophrenics are not differentiated from nonparanoid schizophrenics (McConaghy, 1960; Shakow, 1962; Wynne & Singer, 1963b).

The central symptom in the paranoid schizophrenic which distinguishes him from other Kraepelinian subtypes is the prominence of a delusional system in his thinking. The delusional system may be highly elaborated and systematized, relatively unsystematized, or over time it may become less systematized. Indeed, degree of systematization of delusions and the temporal or chronicity factor appear significantly correlated with differences in attention response patterns among paranoid schizophrenics. In contrast to the paranoid subgroup the predominant symptoms of other schizophrenic subgroups include stupor, negativism, excessive motor activity (catatonia); inappropriate affect, bizarre mannerisms, regressive behaviors (hebephrenia); emotional impoverishment, slow deterioration of behavior, social withdrawal, and fantasy preoccupation (simple schizophrenia).

Process-Reactive Distinction. A second major categorization of schizophrenic patients is based primarily upon developmental and prognostic factors (Becker, 1956; Fine & Zimet, 1959; Garmezy & Rodnick, 1959; Kantor, Wallner, & Winder, 1953). Many laboratory experiments indicate such clarification reduces performance variability between experimental groups and yields signficant differences between the two groups of schizophrenics (e.g., Garmezy & Rodnick, 1959; Herron, 1962; Rodnick & Garmezy, 1957). The process label is applied to patients who exhibit poorly integrated prepsychotic personalities with a continuous and prolonged insidious development of schizophrenic symptoms and who have a poor prognosis. Reactive schizophrenics show a relatively adequate premorbid adjustment, a rapid onset, a stormy clinical picture and have a relatively good prognosis. The terms poor-premorbid adjustment and good-premorbid adjustment are often used interchangeably with process and reactive. The only difference between the two dichotomies stems from an assumption within the process-reactive construct that some kind of organic substrate underlies the pathology of the process group but not that of the reactive group. No such assumption is made in the poor-premorbid–good-premorbid differentiation. However, this does not appear to be relevant to the manner in which subjects are selected for research purposes nor are the various scales which define the two concepts sufficiently different to justify considering them separately.

COGNITIVE CONTROLS OF ATTENTION

Recent studies of individual consistencies in cognitive behavior have emphasized the importance of enduring response dispositions which are relatively invariant over broad ranges of laboratory tasks and environmental conditions (Gardner et al., 1959; Klein, 1958). These response dispositions, termed "cognitive control principles" have been conceptualized as more or less automatic adjustments to the environment which emerge in the course of development and which are shaped by both constitutional and experiential factors. Cognitive control principles refer to various modes of *organizing* percepts, ideas, memories, sensations, etc. These organizational tendencies vary from person to person and reflect typical strategies of coordinating private perceptual and conceptual experiences with those of the "real world."[2] An individual's characteristic pattern of utilizing these organizational tendencies determines the outcome of his coordinations of inner experiences with things and events and constitutes his "cognitive style."

---

[2] For example, some people have an underlying preference for making fine discriminations about nearly everything. This trait may be an asset under certain conditions but it also may prove to be a nuisance in other situations, e.g., the scientist who is always "splitting hairs" both at work *and* at home.

Cognitive controls of attention have been defined in terms of (*a*) subject differences in the extensiveness with which stimuli are sampled when attending to a sensory or perceptual field—the "scanning control," and (*b*) in terms of subject differences in attention to certain segments of complex stimulus fields and simultaneous inhibition of attention to other segments of these fields—the "field-articulation control." A factor analysis of scores on a number of tests of scanning and field articulation, from a large sample of subjects, yielded two orthogonal factors (Gardner, 1961). This finding of independence between the two attention control principles was confirmed in a subsequent study by Gardner and Long (1962a). These investigators also studied various perceptual conditions in which scanning and field-articulation response dispositions codetermined perceptual judgments.

Scanning Control Principle

Piaget (1950), in his discussions of developmental factors related to attention, suggests that early in life perception is subject to much distortion. The young child's attention appears to anchor upon dominant objects in the stimulus field and automatically causes him to overestimate stimuli in the center of the field. These "centration effects" are an inherent feature of the functioning of the perceptual apparatus. In the course of development, the individual ordinarily learns to minimize these misperceptions of apparent size by shifting his attention to and from the center of the perceptual field. In numerous experiments with both children and adults, Piaget and his colleagues have demonstrated that objects in the center of an individual's perceptual field are overestimated. Thus, in certain studies of perceptual constancy children tend to show underconstancy, i.e., to overestimate a near object in relation to a far object because they "center" on the near object. Adults, on the other hand, tend to overestimate the distant object (overconstancy) because they "center" more upon the distant object. The child's judgments tend to be made in terms of the sizes of objects as they are recorded on the retina. In contrast, the normal adult appears to *overcompensate* for this type of error tendency.

Errors in judgment of apparent magnitude have also been found to be a function of the length of time during which subjects "center" upon stimuli in the course of making perceptual judgments (Piaget, Vinh-Bang, & Matalon, 1958). The longer the duration of a centration, i.e., the longer the stimulus is viewed, the greater is the tendency toward overestimation of the stimulus. Errors of overestimation of standard stimuli are regarded as a general tendency to measure less stable parts of the perceptual field (the variable stimuli) in terms of the more stable, the standard stimuli.

Gardner and his colleagues (Gardner, 1961, Gardner et al., 1959, 1960, Gardner & Long, 1962a, 1962b) have elaborated upon Piaget's formulation of centration effects. They have shown that individuals who scan a visual field extensively—who repeatedly look back and forth from one segment of the field to other segments of the field—evidence minimal overestimation or underesti-

---

Another example—some people are readily influenced by pressures from their immediate environments. Others rely primarily on internal cues to guide them in making decisions or in accepting or not accepting popular opinions.

mation of a standard stimulus in certain size estimation procedures. On the other hand, limited scanning behavior is associated with large overestimations of the standard. Precise recordings of scanning behaviors have been made by filming eye movements during size estimation trials (e.g., Gardner & Long, 1962b). In addition, significant individual consistencies leading to the formulation of the "scanning control principle" have been observed in tests of perceptual constancy, incidental noticing, and in a free association test which requires the subject to select associations from internal memory schemata (Gardner et al., 1959). Subjects characterized by extensive scanning responsiveness have also been found to show more Rorschach indices of "isolation" and "projection" as compared with limited scanners (Gardner & Long, 1962b). Finally, studies by Farrow and Santos (1962) and Solley and Santos (1958) indicate that individual differences on the autokinetic illusion and the Necker cube reversible figures procedure may be interpreted in terms of differences in scanning behaviors. These latter studies demonstrate that the scanning response, like any other response, is directly affected by positive or negative reinforcement (also Santos, Farrows, & Haines, 1963).

The scanning mechanism takes on special significance when applied to an analysis of perceptual studies of schizophrenics. Whereas normal groups typically show moderate scanning behaviors on tasks from which scanning responsiveness can be inferred, schizophrenics evidence either marked scanning or minimal scanning (e.g., Harris, 1957). These extreme response dispositions in schizophrenia can be related to the type of symptom pattern, pre-illness history factors, and the stage of the psychotic disturbance. Note that this type of analysis focuses on patterns of constant error scores on the various perceptual tests rather than upon degree of accuracy or correctness. The additional information which can be obtained by this method makes possible new explorations of attention response patterns within and between the various types of schizophrenias.

Perceptual Constancy. Objects of the same size or shape, seen at different distances and at different angles ordinarily tend to be experienced as very similar (perceptual constancy). This phenomenon occurs despite the fact that retinal images produced by two objects viewed at different distances may be quite different. In normal everyday experience and under most laboratory conditions, there is a tendency for overcompensation in the form of overconstancy to occur. Only when spatial and optical cues are markedly reduced or when instructions geared to breaking down constancy perception are used, does the overconstancy tendency diminish in normal, not excessively aroused, and nondrugged subjects (Ardis & Fraser, 1957; Callaway & Dembo, 1958; Carlson, 1962; Holway & Boring, 1941).

In accord with the findings of Gardner et al. (1959), the scanning control mechanism will be used to interpret a number of studies of individual differences in perceptual constancy. As specifically applied to these studies, the relationship between scanning and perceptual constancy is hypothesized to be as follows: The more intensively a subject scans the perceptual field, the greater is the tendency to overconstancy in his perceptual judgments. On the other hand, minimal scanning of the environment is correlated with reduced overconstancy or underconstancy. Thouless (1932) studied the perceptual constancy performances of introverted or "schizothymic" individuals and extraverted or "cyclothymic" individuals. He found that the perceptual performances of schizothymic patients, selected on the basis of Kretchmer's descriptions, tended to be less overconstant than the performances of cyclothymic individuals. This finding of greater size constancy in extraverts than in introverts was confirmed by Weber (1939). Ardis and Fraser (1957) studied shape constancy among subjects who were selected on the basis of scores on the Minnesota TSE Inventory of introversion extraversion. A highly significant relationship was found between introversion as a personality trait and reduced shape constancy.

These perceptual differences between introverts and extraverts fit well with H. M. Voth's (1962) conceptualization of the degree to which such individuals characteristically seek out external stimulation. Voth suggests that the actions and feelings of extraverted-type individuals are determined primarily by external stimulus events ("they live by radar"). The more self-absorbed introverted-type evidence a relative independence from external conditions. The latter group scans the environment to a lesser degree than the former.

Raush (1952) hypothesized that the markedly self-absorbed or socially withdrawn nonparanoid schizophrenic would evidence perceptual underconstancy, whereas the cautious, suspicious, and perceptually vigilant paranoid schizophrenic was expected to show extreme overconstancy. Under ordinary visual conditions, nonschizophrenics, paranoid schizophrenics, and nonparanoid schizophrenics showed overconstancy. The paranoid group, however, was significantly more overconstant than the nonschizophrenic group. When size judgments were made in a darkened room, paranoids evidenced even greater overconstancy judgments than those which were made in the lighted room. In this condition overconstancy accentuation was significantly different from both the nonparanoid group and the control group; the latter two groups tended to reduce overconstancy. Similar findings were obtained by Hartman (1962) in studies of the afterimage observations of paranoid schizophrenics with well-systematized delusions, nonparanoid schizophrenics, and nonpsychiatric subjects. (The phenomenological size of afterimages may be considered as the converse of size constancy since distance cues affect changes in size perception of an afterimage in a reverse manner. The performances of Hartman's experimental groups showed the same kinds of constancy dispositions as Raush's groups, and as predicted in reverse order).

Several studies have reported underconstancy in the size judgments of schizophrenics. Weckowicz (1957) found chronic schizophrenics with hebephrenic and simple schizophrenic syndromes manifested underconstancy when compared with a group of nonschizophrenic mental patients and a group of hospital staff personnel. In a later report (Weckowicz & Blewett, 1959), it was noted that hebephrenics with much thought disorder tended to respond in an underconstant manner whereas paranoids with little thought disorder showed overconstancy (p. 912). Weckowicz, Sommer, and Hall (1958) found that chronic schizophrenics, 17 out of 20 being diagnosed as nonparanoid schizophrenics of various types, showed poor distance constancy. (No performance records were presented). Significant correlations were also found between poor distance constancy and underconstancy in size judgments. Crookes (1957) reported significantly lower perceptual constancy performances in a group of schizophrenics, unspecified as to Kraepelinian subgroup, than in groups of neurotics, psychopaths, and normals.

Lovinger (1956) divided his schizophrenic subjects into "good contact" and "poor contact" groups on the basis of staff personnel ratings. Size constancy perception was studied under three conditions: maximized distance cues (binocular viewing of the stimuli), minimized distance cues (monocular viewing of the stimuli), and no distance cues (monocular viewing, reduction screen, poor lighting). All groups were very similar under conditions of maximum or absent cues. But when minimal cues were offered, poor contact schizophrenics achieved less constancy (overestimation of the closer object) than the good contact and control groups, which did not differ significantly from each other. No significant differences were found between the constancy performances of paranoid and nonparanoid schizophrenics. In attempting to account for the discrepancy between this latter finding and Raush's data, Pearl (1962) stressed the importance of the acute versus the chronic dimension of schizophrenia. He pointed out that Lovinger's "poor contact," underconstant schizophrenics were primarily chronic and that his "good contact," overconstant subjects were mostly early schizophrenics. Similarly, he suggested that Raush's paranoid group were primarily acutes and his nonparanoid group were mainly chronics. Whereas such an inference may be warranted in the Lovinger study, there is little basis for such a statement about Raush's subjects. A more satisfactory explanation of Lovinger's data may rest on the separation of the chronicity variable from Kraepelinian diagnosis.

Autokinesis. Further indications of differences in scanning responsiveness between paranoid and nonparanoid schizophrenics (and between extraverts and introverts) are found in studies of the autokinetic effect (Hahn, 1956; Sexton, 1945; A. C. Voth, 1941, 1947; H. M. Voth, 1962). The autokinetic effect is a visual illusion which usually can be observed by looking at a stationary point of light, from a short distance, and for a period of time, in an otherwise dark room. Marked differences in amount of perceived movement in this situation have been found between introverts and extraverts and between different psychiatric groups. Introverted subjects perceive more movement than extraverted subjects and, within the schizophrenia classification, nonparanoid subgroups report much greater movement than paranoid schizophrenics. These findings may be viewed as an extension of the "centration effects" formulation. Piaget and his colleagues had demonstrated that differences in the perceived magnitude of a stimulus are a function of the number of repeated "centerings" upon the stimulus and also of the length of time during which subjects "center" upon the stimulus. It appears that in the autokinetic situation subjects who scan extensively, i.e., who repeatedly "center" on and away from the light point, evidence relatively small amounts of perceived movement. On the other hand, minimal scanning, i.e., minimal "centering" and excessive fixation upon the light point, is correlated with significantly greater amounts of perceived movement.

A. C. Voth (1947) reported a study in which groups of catatonic, hebephrenic, and simple schizophrenic patients showed much perceived movement of the light in the autokinetic situation; individuals diagnosed paranoid schizophrenic, true paranoia, paranoid condition, and involutional paranoia reported little movement. It appeared that the more prominent paranoia, without bizarre ideation, was in the symptom pattern, the greater was the tendency to report no autokinetic movement. *It was found that either extreme movement or no movement at all were poor prognostic signs.*

Sexton (1945), in his autokinetic study of schizophrenics, found that 36 out of 38 nonparanoids (18 catatonic and 20 hebephrenic) evidenced a "large" or "maximum" amount of movement whereas 7 out of 8 paranoid schizophrenics evidenced lesser amounts of movement. He also reported that the greater the degree of perceived movement, the greater was the extent of clinically observable withdrawal behavior. Hahn's (1956) findings were consistent with this and also indicated large amounts of perceived movement among nonparanoid schizophrenics. In Hahn's experiment, chronic nonparanoids were assigned to one of three reinforcement conditions—reward, punishment, and ordinary testing conditions. They were tested individually in the autokinetic situation before and after the administration of success, failure, and control treatments. Success experiences were found to decrease perceived movement and failure experiences were found to increase perceived movement. Such data lend support to the hypothesis that changes in the degree and direction of scanning of the environment are correlated with shifts in attention which have approach and avoidance utility (Santos, Farrow, & Haines, 1963).[3] These results further suggest that nonparanoid schizophrenics evidence more than minimal scanning only when such behavior is followed by self-satisfaction and a reduction of anxiety. They will evidence minimal scanning, resulting in constriction of the perceptual field, when attending to the field is followed by punishment and an increase in anxiety. The generalizing of this aspect of the attention response, i.e., the formation of a minimal or excessive scanning response disposition, is one way of conceptualizing the approach-avoidance behaviors of all organisms (Mowrer, 1960). It is also in line with Solley's and Murphy's (1960) and Guthrie's (1959) contentions that the attention response, like any other response, may be conditioned.

One last point needs to be made here. It is extremely important to distinguish scanning behaviors which are shaped in and are specific to a given experimental situation from characterological scanning dispositions. It is not difficult to alter the attention behaviors of schizophrenics who are either maximum or minimum scanners by controlling specific reinforcement contingencies in the experimental situation (Silverman, 1963a). However, the more enduring aspects of a scanning response disposition are apparently based upon a long and stable history of reinforcement contingencies and are highly resistant to change.

Size Estimation. Filmed eye-movement recordings of scanning behaviors have indicated that individual differences in scanning behavior can be reliably inferred from a particular form of size estimation procedure (Gardner & Long, 1962b). This size estimation procedure was employed in a recent study of schizophrenic patients in which marked scanning differences were hypothesized between paranoid and nonparanoid groups (Silverman, 1964a). In accord with the Gardner et al. (1959) formulation, the presumed minimal scanning nonparanoids were expected to evidence large overestimations of the standard stimuli, whereas extensive scanning paranoids were expected to underestimate stimuli. The patients, whose diagnoses were unknown by the experimenter, were instructed to adjust a circular patch of light until it looked exactly like the size of a hand-held disk. Three different sized disks were presented four times each, in random order, for a total series of 12 judgments. Analyses of the mean size estimation differences between diagnostic groups yielded significant *p* values ranging from .0025 for the three separate stimuli to .0006 for the total series of judgments. These marked differences were consistent with Piaget's centration hypothesis and the Gardner et al. scanning hypothesis.

---

[3] Alterations in the degree and direction of scanning responsiveness have also been achieved among normal subjects by differentially reinforcing perceived movement in different regions of an autokinetic field (Farrow & Santos, 1962).

The application of this type of an analysis to some size estimation studies of poor-premorbid and good-premorbid schizophrenics indicates significant scanning differences between the two groups. Further, these studies highlight the effect of anxiety level, physiological arousal, and degree of chronicity. (It will be recalled that the poor-premorbid classification describes individuals who show gradual or insidious prodromal signs of schizophrenia beginning relatively early in life as contrasted with the developmentally more advanced good-premorbid schizophrenics with a good prognosis who show a relative absence of early prodromal signs and a rapid, stormy onset).

Harris (1957) studied the size estimation performances of poor-premorbid and good-premorbid schizophrenics, (divided on the basis of the Phillips Prognostic Scale) and a nonpsychiatric group. The subjects were asked to judge from memory the sizes of previously presented pictures which had been selected for their presumed cue relevance for the patients. These pictures included mother-son scenes of overprotection, rejection (ignoring), domination, and feeding (acceptance), together with a more neutral tree-bush scene and a geometric figure (square). Significant differences were found between the two patient groups which were not at all apparent when the total schizophrenic group was compared with the nonpsychiatric group. The poor-premorbid group tended to judge the stimuli to be larger than they actually were (minimal scanning); the good-premorbid group tended to underestimate them (extensive scanning); and the nonschizophrenic group showed minimal errors in size estimation ("balanced" scanning). Furthermore, the overestimation or underestimation tendencies were also apparent when the frame containing the neutral picture was judged (Fig. 1).

In a subsequent study by Zahn (1959), positive and negative affective value were first conditioned to neutral pictures during a discrimination task not involving differential size judgments. Subjects' responses to one of the pictures were rewarded by the comment "RIGHT" and their responses to the other picture were punished by the comment "WRONG." A pilot study, in which subsequent judgments of the sizes of the stimuli were made from memory, was consistent with that of Harris. In addition to a general overestimation-minimal scanning tendency on the part of the poor-premorbid group, Zahn found that the poor-premorbid group overestimated the size of the picture associated with punishment relative to their size estimations of the rewarded picture. Good-premorbid subjects showed tendencies in the opposite direction. The interaction between premorbid adjustment and reactiveness to type of reinforcement was statistically significant.

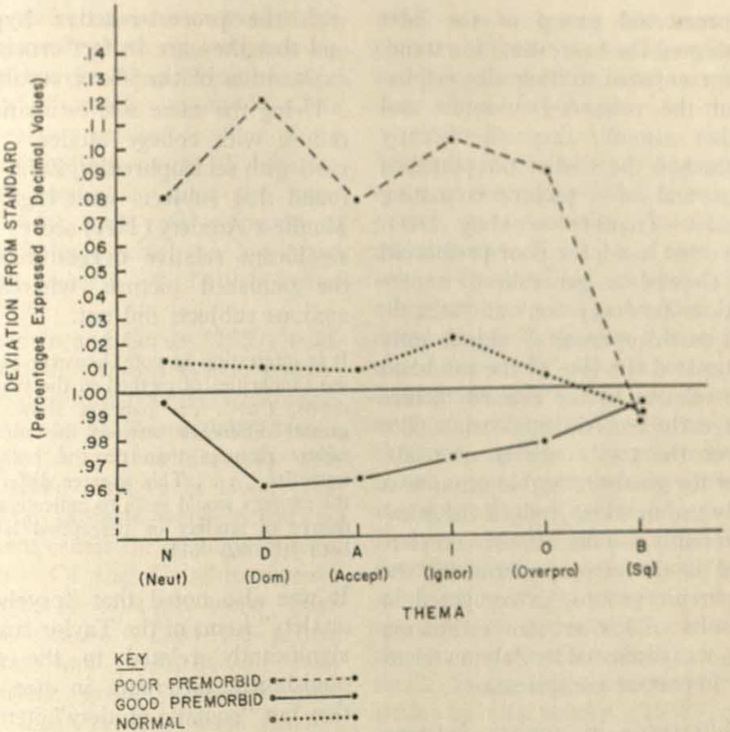


FIG. 1. A comparison of mean size estimates of good and poor premorbid schizophrenic and normal groups for all scenes. (From Harris, 1957).

Both of these size estimation studies indicate the presence of a correlation between extreme scanning response dispositions and the good premorbid-poor premorbid or process-reactive classification. These experiments also support the contention that premorbid history factors are related to genetic levels of perceptual development (Becker, 1956; Fine & Zimet, 1959). Thus, the developmentally earlier appearing minimal scanning response (Piaget, 1950) was characteristic of the poor-premorbid group, especially in their estimations of the pictures associated with punishment. The good-premorbid schizophrenics tended to react to these same pictures by extensively scanning them and thus reducing their apparent magnitude. Note here that the scanning-response construct also provides a heretofore elusive explanation for the translation of attitudes toward pleasantness and unpleasantness into changes in estimated size. For example, individuals who ordinarily evidence minimal scanning, react to unpleasant stimuli by further reducing their scanning responsiveness and thus enhancing the perceived magnitude of stimuli. Underestimating-extensive scanners appear to scan to an even greater degree when confronted with unpleasant stimuli.

That the relationship between premorbid adjustment and size estimation performance is far more complex than was first thought is illustrated in a later study by Zahn. In contrast to the results of the earlier studies, the good premorbid group of the later study showed the overestimation trend. That is, compared to their size estimations of the rewarded stimulus and the other stimuli, they significantly overestimated the size of the punished stimulus and of a picture connoting domination (mother scolding son). On the other hand, the poor-premorbid group showed a generalized underestimation tendency on all stimuli. The nonschizophrenics significantly overestimated the size of the punished picture relative to the reward picture on the early experimental trials, but not over the total number of trials. Further the good-premorbid group was the only group which judged the stimuli differently. This almost complete reversal in the error patterns for the schizophrenic groups, as compared to the results of earlier size estimation studies, was discussed by Zahn in terms of two important considerations:

1. Differences in anxiety between the schizophrenic groups may have affected size judgment performances. What qualitatively had appeared as higher anxiety in the good-premorbid group could have somehow operated to produce overestimation of the picture associated with failure (punishment).
2. Significant differences in the length of hospitalization existed between subjects who took part in Zahn's latter study and subjects who participated in Harris' study and Zahn's pilot study. Whereas mean lengths of time since first psychiatric hospitalization for the "goods" and "poors" in Zahn's latter study were 6.00 years and 8.09 years, respectively, in both of the earlier studies these means were 3 years and less.

A careful search of the literature indicates that, indeed, both level of anxiety or emotionality and degree of chronicity are inextricably interwoven with the process-reactive hypothesis and that they are in fact crucial to an explanation of the above results.

Using the same size estimation procedure with college students as was used with schizophrenics, Zahn (1960) found that subjects with high Taylor Manifest Anxiety (HA) scores showed significant relative overestimations of the punished picture, whereas low-anxious subjects did not.

> It is interesting to note, however in view of the similarities of method in the two experiments, that . . . patients show a much greater difference between the success and failure pictures than do the present HA normals. . . . . This greater difference for the patients would seem to indicate a greater degree of anxiety or perceptual lability or both [pp. 460–461].

It was also noted that "psychological anxiety" items of the Taylor scale were significantly related to the reward-punishment difference in size estimation but "somatic anxiety" items were not.

Ilanit (1961), using the Elgin Prognostic Scale to divide patients into process and reactive groups, took special care to restrict his population to acute schizophrenics. Patients selected had no history of previous psychiatric hospitalizations and had been in the hospital 18 months or less. Two measures of anxiety, the Taylor Manifest Anxiety (*MA*) scale and the Welsh Anxiety Index, were administered to all subjects along with various other psychological tests. Whereas no differences were found between the process and reactive groups on the *MA* scale, the Welsh Anxiety Index revealed significant differences in a direction opposite from that which would ordinarily be expected and from that predicted by Ilanit. The early process group had significantly higher anxiety scores than did the early reactive group. The failure of the *MA* scale to differentiate between groups may very well prove related to a distinction between "psychological anxiety" and "somatic anxiety." This conjecture is presented in view of the significant findings on the Welsh anxiety scale especially since the items on the Welsh scale refer primarily to personal feelings which an individual has about himself, i.e., "psychological anxiety."

Zuckerman and Grosz (1959) studied autonomic nervous system reactivity in process and reactive schizophrenics by means of the mecholyl test, a cholinergic drug that normally produces marked changes in blood pressure. The subjects were patients at a treatment center for acute psychiatric disorders. Of the 27 schizophrenics tested, only 4 of them were "post-treatment" cases. Thus the patient population was primarily an early schizophrenic group. The subjects, divided into process and reactive groups on the basis of the Kantor, Wallner, and Winder (1953) criteria, were administered mecholyl on four separate occasions since previous research (Grosz & Miller, 1958) had indicated considerable day to day variability in response to this drug. The early process group showed a significantly greater drop in systolic blood pressure following injection of mecholyl as compared with the early reactive group.

This finding is in direct contrast to reports by King (1958), Funkenstein, Greenblatt, and Solomon (1950), Meadow and Funkenstein (1952) and others. In these latter studies all schizophrenic subjects had been hospitalized on at least one occasion. Furthermore, the results of similar investigations with long-term chronic schizophrenics (e.g., Meadow, Greenblatt, Funkenstein, & Solomon, 1953; Bockoven, Greenblatt, & Solomon, 1953) would suggest that most of the patients in these "contradictory" studies had long amounts of total hospitalization. Data on number of previous hospitalizations, time elapsed since first hospitalization, and total length of time spent in mental hospitals were not reported in any of the "contradictory" research reports. In any case the results of mecholyl test studies of chronic schizophrenics indicate that strong drops in blood pressure are not associated with the process category as Zuckerman and Grosz found in their study of acute patients. Rather among chronics, reactive schizophrenics with initial good prognoses show the stronger drops in blood pressure. Process chronic schizophrenics evidence minimal drops in blood pressure following mecholyl infusion. Such responsiveness appears to occur *after* they have been schizophrenic for some time. Supporting this contention is a study by Weckowicz (1958) of the relationship between response to mecholyl and perceptual constancy among twenty long-term chronic schizophrenics (8 years mean hospitalization). Subjects tending to show normal overconstancy had a blood pressure response pattern most similar to one associated with poor prognosis (process schizophrenia) and least similar to one associated with good prognosis (reactive schizophrenia). That is, blood pressure changes for the overconstant perceivers corresponded to the Group I-poor prognosis pattern found by Funkenstein and his colleagues, i.e., an initial drop followed by a larger rise in pressure. Underconstant subjects evidenced lesser or no subsequent rises in blood pressure.[4] Such data are con-

[4] In addition, Weckowicz found that weaker drops in blood pressure tended to be associated with overconstancy. That this correlation did not reach statistical significance was probably due to a more constricted

TABLE 1

INFERRED RELATIONSHIPS BETWEEN PERCEPTUAL, EMOTIONAL, AND AUTONOMIC VARIABLES IN PROCESS AND REACTIVE SCHIZOPHRENICS

	Mean length of all hospitalizations	
Group	0–3 years	6 years and over (approximate)
Process	Size overestimation-minimal scanning	Overconstancy and size underestimation-extensive scanning
	Higher anxiety	Lower anxiety (inference)
	Stronger drops in systolic blood pressure following mecholyl	Weaker drops in systolic blood pressure following mecholyl
Reactive	Size underestimation-extensive scanning	Underconstancy and size overestimation-minimal scanning
	Lower anxiety	Higher anxiety (inference)
	Weaker drops in systolic blood pressure following mecholyl	Stronger drops in systolic blood pressure following mecholyl

sistent with Zahn's findings with chronic poor-premorbid and good-premorbid patients wherein the poor-premorbids and not the good-premorbids evidenced perceptual performances most similar to those of nonschizophrenics.

Taken together these data indicate that much of the apparent contradiction in the literature may be due to the confounding effects of varying chronicity within groups of patients who otherwise are selected in a reasonably careful manner. On the basis of studies which have controlled for degree of chronicity, either inadvertantly or otherwise, it is suggested that early process schizophrenics manifest emotional, autonomic, and attention response characteristics almost completely opposite to those of chronic process patients. ' Opposite changes also have been found between early and long-term chronic-reactive schizophrenics (Table 1).

Field-Articulation Principle

The previous section of this paper reviewed and discussed studies on the breadth of perceptual scanning under relatively free conditions. The present section will be concerned with the selective aspects of attention, especially those response mechanisms which determine what elements in a stimulus field exert dominant influence over behavior. Studies by Piaget (1950) and his colleagues, Witkin and his colleagues (Witkin, Dyk, Faterson, Goodenough, & Karp, 1962; Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954), and Wapner and Werner (1957) have demonstrated that perceptual differentiation of complex stimulus fields evolves in the course of development. The experiences of one's body, self, and the external environment which early in life are essentially

---

range of values which are associated with drops in blood pressure as compared to rises. Thus, Funkenstein et al. (1950), reported that, for schizophrenics, blood pressure response patterns following mecholyl tend to cluster to a greater extent in two of seven groups, both of which are characterized by minimal drops in pressure. In Weckowicz's sample of long-term chronic patients, unselected as to process or reactive type, it is quite probable that subjects clustered primarily into the two minimal drop-in-pressure groups.

global, vague, and diffuse, become progressively more articulated. However, even among "normal" adults wide variations have been found in the degree of clarity with which the self and external objects are segregated from the total perceptual field. These variations have been observed in laboratory procedures requiring: (*a*) the perceptual separation of objects from their surrounds such as in the rod-and-frame test, (*b*) the differentiation of field cues from sensory cues such as in the tilting-room–tilting-chair task, (*c*) the perceptual restructuring of a stimulus field such as in the embedded figures task, and (*d*) the perceptual inhibition of distracting, intrusive cues such as in the color-word test or on the Muller-Lyer illusion.

The concept of field articulation was elaborated by Gardner et al. (Gardner, 1961; Gardner, Holzman, Klein, Linton, & Spence, 1959; Gardner, Jackson, & Messick, 1960) to explain individual differences on tasks of "perceived incongruity" in terms of differences in the selectiveness of attention. It was pointed out that on all of the above mentioned procedures, effective performance was dependent upon directing attention to one of two sets of cues that induced opposing response tendencies. The field-articulation principle thus encompassed Witkin's concept of field dependence-independence (Witkin et al., 1954), Klein's constricted-flexible control principle (Klein, 1954) and aspects of Piaget's "strategies" of attention formulation (Piaget, 1950). Factor analyses of subjects' performance scores indicated significant intercorrelations among the tasks which defined these various concepts. Constant error performance patterns were consistent and reliable.

In a recent study by Weiss and Sherman (1961), the Stroop color-word test scores of acute and chronic paranoid and nonparanoid schizophrenics were correlated with their scores on the Taylor *MA* scale. The color-word test requires that subjects not attend to the printed names of colors but that they restrict their attention to rapidly naming the colors in which the words are written. The initial hypothesis tested was that high anxiety during a task which evokes competing response tendencies would result in greater response interference and thus poorer performance. Low anxiety scores on the *MA* scale were expected to be associated with more efficient performances. The predicted relationship between anxiety level and interference was found for chronic schizophrenics (median total hospitalization, 8 years) but not for acute schizophrenics (median total hospitalization, 9 months). Further analysis indicated that the more anxious subjects in the acute paranoid schizophrenic category actually tended toward *decreased* response interference on the color-word test. The difference between the *MA* scale-Stroop correlations for acute paranoids and chronic paranoids was highly significant. No differences were found between the acute and chronic patients without paranoid features.

These results suggest that the paranoid's readiness to deal analytically with a complex stimulus field may have a characterological aspect which is readily observable in the acute stages of the disorder. Furthermore, this disposition is most apparent under conditions of moderate or high anxiety. The finding that chronic paranoids did not employ this type of responsiveness when extraordinarily anxious could indicate that the field-articulation response disposition is significantly affected by chronicity per se. Additional evidence in support of this latter hypothesis will be presented at a later point.

Witkin et al. (1954) reported data on the tendency of hospitalized mental patients to appear at the extremes of the distribution, when compared to nonpsychiatric patients, in their performances on the rod-and-frame test and the tilting-room–tilting-chair test. The psychiatric patients showed a striking concentration of cases in either the extreme field-independent or the extreme field-dependent groups. Only a small proportion of hospital cases were found in the intermediate performance group. This was in contrast to the nonpsychiatric subjects who tended to fall mainly into the intermediate performance category and rarely in either of the extreme categories. Out of 12 males diagnosed as paranoid schizophrenic, 7 were found to be markedly field-independent (high field-articulators), 2 evidenced intermediate category performances, and 3 were markedly field-dependent (low field-articulators). No data on male nonparanoid schizophrenics were reported. The fact that female schizophrenics, regardless of diagnostic subtype, tended to be markedly field-dependent was consistent with the finding that women, as a group, were considerably more field-dependent than men.

Taylor (1956) studied the embedded figures test (EFT) performances of schizophrenics who were rated as being primarily delusional or primarily hallucinatory. In this task, the subject is asked to find a particular simple figure within a larger complex figure. The simple figure is hidden by being part of the pattern of the more complex figure (e.g., the outline of the simple figure may partly define the boundaries of several more prominent patterns in the complex figure). In the Witkin studies, subjects who performed in a field-dependent manner on the rod-and-frame and tilting-room–tilting-chair tasks took a significantly longer time than field-independent subjects to isolate the embedded figures from their surrounds. In Taylor's experiment, delusional and hallucinatory subjects were selected on the basis of ratings on a series of items derived from the Lorr scale for rating psychiatric patients. Schizophrenics who exhibited both delusional and hallucinatory symptoms were excluded from the study. Results were consistent with the hypothesis that delusional subjects are more field-independent than hallucinatory subjects as measured by their EFT performances.

Witkin et al. (1962), recently reviewed a number of studies of field-dependence–field-independence among various clinical groups. On the basis of the evidence considered, the authors concluded that,

> when they become psychologically ill, persons with contrasting modes of field approach are apt to develop different kinds of pathology [p. 209].

Some theoretical speculations were also offered about the etiology of the field-independent approach of patients in whom paranoid (delusional) reactions form a predominant symptom. In accord with the formulations of Freud (1958), Cameron (1947), and others, Witkin et al. (1962) suggested that the paranoid structures his environment by projecting his own system of ideas upon it. He does so in a highly selective fashion, and is especially prone to implicate particular people and specific situations.

> Such selectivity requires that experience of the world be articulated. In this connection, the paranoid is noted for the detailed, articulated system he develops. . . . The use of projection as a device for self-preservation, however bizarre, presupposes some degree of differentiation [p. 209].

From this it would follow that the more perceptually and conceptually articulated the manner in which the paranoid experiences his world, the more systematized will be his delusional complex of ideas.

Another implication to be derived from the field-articulation construct is suggested by some perceptual studies of children (Piaget, 1950; Wilkin et al., 1962). These studies indicate that the disposition to either a highly differentiated or poorly differentiated field-articulation response is present rather early in the course of development. In fact, studies by Escalona and Heider (1959) and others would suggest that these dispositions are present in the earliest days of life. It is conceivable then that such dispositions may be significant determinants of the type of schizophrenia that an individual develops. One would therefore expect differences in the pre-illness adjustments of schizophrenics to be correlated with degree of field articulation. Thus, Bryant (1961) divided chronic schizophrenics into poor-premorbid (process) and good-premorbid (reactive) groups and studied their performances on the embedded figures test and on the rod-and-frame test. Subjects for each group were selected from those with extreme scores on the Phillips Premorbid Adjustment Scale. It was hypothesized that the reactive group would perform on these tests in an analytic, field-independent manner and that the process group would evidence field-dependent performance. Significant differences in group performances were found in the predicted directions. Mean deviations of the reactive group on the rod-and-frame test approximated the mean obtained by a college group in the Witkin et al. (1954) study. On the other hand, the mean deviation of the process group approximated the mean deviation which Witkin et al. found for 8-year-old children.

These findings become even more meaningful when considered in relation to personality factors which Witkin et al. found to be associated with field-dependent and field-independent performances. Thus field-dependent individuals typically evidenced such attributes as: an inability to function independently of environmental supports, an absence of initiating activity, low self-esteem, difficulty in controlling anxiety, and a relatively primitive, undifferentiated body image. The description of the process (field-dependent) patients in these terms is consistent with the data on their inadequate social-sexual adjustments (Garmezy & Rodnick, 1959). By the same token, the personality correlates of field independence are appropriate descriptions of the reactive patients who evidence more adequate social-sexual premorbid adjustments. Field-independent perceivers tend to be characterized by activity and independence in relation to the environment, relatively higher self-esteem, more effective techniques for controlling anxiety, and a more differentiated body image. Note here that field dependence–field independence and adequacy of premorbid adjustment are not related to presence or absence of schizophrenia. Rather these data indicate that schizophrenia assumes different forms among people with relatively analytic, field-independent, and relatively diffuse, field-dependent orientations (Witkin et al., 1962).

The differences in perceptual orientations of process and reactive schizophrenics also are consistent with the conceptualization that premorbid history factors are related to genetic levels of perceptual development. In accord with the developmental levels hypotheses of Werner (1948) and Piaget (1954), the extreme perceptual orientations of process and reactive schizophrenics are interpreted as representing lower and higher levels, respectively, of sensory, perceptual, and conceptual differentiation (Bryant, 1961).

DEFENSIVE ASPECTS OF SCANNING AND FIELD ARTICULATION

Bleuler in his writings on the problem of attention in schizophrenia made the following comments:

> If "enslavement by optical impressions" is a pathological overemphasis of the sense impressions, its opposite is that under other circumstances the patient ignores the outside world completely.
>
> Between these two extremes, all transitions are possible. . . . At the one extreme patients are easily distractible, seem to lack directives of their own and depend entirely on outer impressions; this is the case when the symptom of naming prevails; at the other extreme they are undistractible and even the strongest stimuli will fail to influence their train of ideas or arouse their attention.
>
> In their daily life the majority of chronic cases show little that is remarkable in this respect: they are at their work and attend to the new event insofar as it interests them. They can be interrupted in their talk. In more severe cases, however, the distractibility is diminished. . . .
>
> Lack of distractibility may appear to spring from the patients' *lack of interest:* as they care about nothing, nothing can influence their behavior. But it can be shown that the same patients understand very well what goes on about them, even when they pay no active attention. In certain acute conditions distractibility is greatly reduced, often even to the vanishing point. [Rapaport, 1951, p. 640].

Attention-response dispositions emerge in the course of physical and psychological development. Like physiological and psychological needs, they are directive; they exert a selective influence on the cognitive field (Klein, 1954). Unlike needs, however, they are not necessarily linked to a class of satisfaction-giving objects. Rather they are activated by and operate upon perceptual and conceptual incongruities (e.g., comprehending an ambiguous or complex stimulus field). Although conceptualized as learned behaviors, they are not simply "attentional habits" which are correlated with certain qualities of stimuli. Rather they are conceived of as general response dispositions which may express themselves in broadly equivalent ranges of environmental situations. It should be noted that since the two attention control principles being discussed here may operate independently of one another, various combinations of the two controls are likely to occur in an individual (e.g., high scanning responsiveness-low field articulation responsiveness, etc. The diagnostic and behavioral implications of these various combinations of controls for the schizophrenias are now being investigated Silverman, 1963b).

The concept of "defense" is extremely useful when relating attention response patterns to modes of mastering anxiety. Thus it is suggested that early in the development of a schizophrenic disorder, as isolations, repressions, and denials become less and less effective, the individual begins to rely on the most basic forms of adient-abient adjustive mechanisms. These mechanisms include formerly "conflict free" attention-response dispositions. In this respect, schizophrenics resemble very young children whose earliest defensive behaviors are ones which involve motivated attention. Solley and Murphy (1960) have suggested that among the earliest aspects of the environment that are attended to are those complex stimulus sources necessary for biological survival. These stimulus sources include internally produced and environmentally produced anxiety cues which serve as discriminative stimuli for the activation of attentional responses. External psychologically noxious cues which are consistently coincident with attentional responses also come to elicit these response dispositions (excessive or minimal) in their own right. Under conditions of severe psychological stress, the characteristic mode of attention responsiveness utilized by the preschizophrenic is exaggerated and takes on a singularly defensive function. That is either via what Bleuler referred to as "enslavement by optical impressions" or by "completely ignoring the outer world," the schizophrenic gains extraordinary control over his perceptions. (As William James noted long ago, the mere reception of a stimulus is insufficient for perception since an individual's experiences depend upon what he agrees to attend to.)

Consider, for example the schizophrenic who relies on excessive scanning and who is constantly preoccupied with this orienting behavior. The development of this type of disposition is referrable to a particular set of environmental contingencies in which excessive scanning has consistently led to an attenuation in the intensity of anxiety. The excessive scanner has learned that his most effective means of escaping or avoiding anxiety is to be hyperalert to the presence of cues which often precede or co-occur with noxious events (Berlyne, 1960). However, while such behavior serves to maximize the receipt of information about impending "threatening" events, it also has the paradoxical effect of increasing the probability that sources of aversive stimulation will be encountered. Under these conditions an excessive scanning disposition usually becomes coordinated with a propensity for constant reinterpretations of relations between events and the motives of others in a desperate attempt at protecting one's self-esteem.

Research by Margaret Singer and Lyman Wynne (1963; Wynne & Singer, 1963a, 1963b) and clinical observations by Arieti (1955) suggest that an excessive scanning response may be fostered rather early in life by an unusual form of communication between parents and their preschizophrenic children. Singer and Wynne have observed that schizophrenics with "fragmented" thought disorders,[5] showing the "cognitive alertness" of high scanners, have parents with particular kinds of communication problems. On their Rorschachs and TATs these parents evidence tangential ideas, distracting comments, and rather extreme non sequiturs. Their TAT stories are replete with fragments of alternative meanings and alternative feelings, usually none of which is selected out and completed. Furthermore, they tend to build an attitudinal context into their communications, in which the sources of ideas appear to originate outside themselves. In turn, their children respond in cryptic, paranoid-like ways, e.g., they unearth hidden meanings in events and objects, perceive in terms of odd vantage points, treat new stimuli as familiar, and often experience their thoughts as not their own. In contrast to patients whose early histories indicate a poor-premorbid social adjustment and a different family communication pattern, schizophrenics evidencing a "fragmented" type of thought disorder usually have had distinctly better premorbid social adjustments.

Arieti (1955) comments on the early social history of the excessive-scanning paranoid schizophrenic as follows:

---

[5] The term "fragmented" is applied to forms of schizophrenic thought disorder in which relatively differentiated styles of attending and perceiving are apparent but wherein there are serious failures in the capacity to integrate ideation with impulse or affect and events with their contexts. This type of thinking disorder is found quite often among paranoid schizophrenics. Schizophrenics with "amorphous" forms of thought disorder evidence predominately undifferentiated attentional and perceptual responsiveness (i.e., minimal field articulation).

> The parents of the paranoid type do not criticize the patients for their actions; they generally *accuse* the patients for their *intentions* or for *lying*. The child learns to defend himself by anticipating these accusations, and therefore becoming anxious and suspicious, or by developing a faculty for rationalizations. He has to find almost a legal or technical way to protect himself from insinuations and accusations [p. 139]. . . . Other cases instead of focusing their attention on their pseudorational defenses, seem to sense or magnify any kind of hostility in the environment [p. 140].

The defensive maneuvers of the preparanoid thus dispose him to constantly scan environmental input for possible threats to self-esteem and to deal with such threats by selectively examining and translating their meanings.

Additional observations by Singer and Wynne (1963) and by Arieti (1955) indicate that early social interaction factors may foster the development of minimal scanning and/or minimal field-articulation dispositions. The parents of schizophrenics with "amorphous" thought disorders communicate with their offspring in an exceptionally vague or pseudospecific manner. There is no effort at clarifying obscure perceptual-associative responses and the child's emotionally disturbing experiences are dealt with in vague, global ways or not responded to at all. Thus environmental avoidance, blurred, and uncertain perception, and poorly organized ideas are all subtly reinforced, whereas responses geared toward perceptual and conceptual differentiation most often result in increases in anxiety.

In a discussion of the social history peculiarities of nonparanoid schizophrenics, Arieti (1955) suggests that more so than in other types of schizophrenia, catatonic schizophrenics, as children

> are unable to accept their environmental conditions, and at the same time they are unable to fight them [p. 125].

Their break with reality occurs through "withdrawal from action." Such behavior is apparently based upon their low subjective probability estimate that consensually valid overt actions reduce or terminate anxiety.

> The hebephrenic type occurs mostly under two sets of circumstances: either when an arrest at a paranoid level does not remove anxiety or when the environmental factors do not predispose to the development of paranoid defenses [pp. 155–156].
>
> Frequently in the development of the simple type of schizophrenia, in addition to poor child parent relationships . . . the patient has had to compete with a younger sibling. . . . The patient abandons the fight . . . and refuses to grow psychologically [p. 156]. Rather than change reality, the simple schizophrenic limits reality. He narrows his horizons to a large extent, so that he will be able to make some compromise with what is left of reality [Arieti, 1955, pp. 187–188].

Minimal-scanning and minimal field-articulating schizophrenics appear to avoid anxiety by directing their attention away from the environment and onto internal processes such as hallucinations. Yet if this were completely the case, such individuals should experience even greater anxiety than that which originally motivated their withdrawal. That is, the chaotic nature of a prolonged, self-induced sensory deprivation state would operate to prevent such schizophrenics from completely removing themselves from their environments. They must register stimulation in order to live. Rapaport (1960) cites evidence for the position that the simplest everyday behaviors of waking life are developed and maintained only by "reaching out for stimulation." He supports his position using results of the isolation experiments by Bexton, Heron, and Scott (1954) and others; the hypnosis studies by

Brenman and Gill (1947); and the sensory disarrangement experiments by Werner and Wapner (1955) and others. Of particular interest are a series of experiments by Held (e.g., Held & Hein, 1958) which indicate that our usual spatial orientations, however stable they may seem to us, are in the normal course of life brought about and maintained principally by "*re*afferent stimulation." It seems clear then, that what may appear in schizophrenia as the directing of attention away from the environment cannot be considered as a total shift of attention onto internal processes. Rather what appears to be involved are: (*a*) an anchoring of attention on dominant objects in the stimulus field (minimal scanning) and/or (*b*) global and unarticulated attentiveness to sensory inputs. Such fixedly motivated attending operates selectively to minimize or "gate out" much of the perceptual and conceptual input from other sources. The neurophysiological basis for such a gating mechanism is well documented in the research literature for both animals and humans (Berlyne, 1960, Ch. 3).

Clinical observations by Arieti (1961, 1962), Rapaport (1951), and Bleuler (1951) support such a conception of environmental withdrawal in schizophrenia. Arieti (1961) has frequently found that very regressed or very acutely ill patients perceive integrated stimulus patterns, not as unities, but rather in an extraordinarily unintegrated manner.

> For instance, a patient looking at the nurse could not see or focus on her as a person but perceived only her left or right eye or her hand or her nose, etc. Another patient who, while she was in a state of dangerous excitement was put in a seclusion cell, remembered that she could not look at the whole door of the cell. She could see only the knob or the keyhole or some corner of the door [p. 8].

Arieti also observed that with increases in excitement, these perceptual subunits divide into even smaller fragments. The disintegration of wholes phenomenon, termed "Awholism," apparently is experienced by the schizophrenic as *happening to him* and he may even attempt to reconstruct his fragmented wholes into unities. When this occurs the reconstructed wholes turn out to be unlike the original ones and a different type of perceptual distortion results.

Rapaport's (1951) explanation of hypermnesia also lends itself to the position presented here. (Hypermnesia refers to the capacity of the schizophrenic to recall experiences as well or better than normals while making no differentiations between the salient and irrelevant aspects of these experiences). Rapaport states that although attention may be temporarily withdrawn from reality, the schizophrenic invariably makes a new attempt at contact with his environment. Such new contact is

> appreciably facilitated by encountering reality piecemeal, so as to guarantee that no relationships will be met which would arouse impulses and anxieties. It is noteworthy that recall of meaningful material fares very badly in these bland (chronic schizophrenic) conditions [p. 628].

Bleuler (1951) also has noted that,

> even though uninterested and autistically encapsulated patients pay little attention to the outside world, they register a remarkable number of events of no concern to them. The selection which attention exercises over normal sensory impressions may be reduced to zero, so that almost everything that meets the senses is registered [p. 56].

Similar behaviors have been recorded by Kraepelin (1919).

The above clinical observations, as well as the researches presented earlier, indicate that the notion that attention is directed away from the environment in schizophrenia is in need of major reformulation. The attention-response formulation presented in this paper is suggested as a general interpretative statement of the clinical and experimental literature on disturbances of attention in schizophrenia. Thus Arieti's observation which he termed Awholism (the inability to perceive wholes) may be interpreted as being due to extreme minimal scanning or fixation of momentarily dominant objects in the visual field—analagous to pathological staring behavior observed in people who are in semistuporous conditions. Rapaport's and Bleuler's observations on adult schizophrenics lend themselves quite well to an explanation in terms of pathologically low field articulation such as was found among young children by Witkin and which usually becomes less and less pronounced with age. Furthermore, it just does not make sense to say without specific qualification that hypermnesic people pay little attention to the world around them. If in fact one active perceptual event inhibits all other perceptual acts depending upon their respective degree of incompatibility (e.g., direction of the perceptual act) how can someone both attend to the outside world "hypermnesically" and still be said to live in an autistic encapsulated world.

On the other hand, an explanation phrased in terms of exaggerated attention-response mechanisms has considerable explanatory power. These mechanisms are conceptualized in terms of an interaction of essentially two independent processes, scanning and field articulation, which in particular (and as yet unknown) combinations predispose schizophrenics to structure their environments in different ways.

A Schizophrenic Defense Mechanism

At this point there still remains the task of tying together the attention-response conceptualization with two important developmental aspects of schizophrenia—(*a*) the historical variables which have contributed to the "choice" of the schizophrenic resolution and (*b*) the effects of long-term immersion in the schizophrenic psychological environment. In regard to the first of these issues, there is impressive evidence which suggests the importance of early experiences on the development of a particular kind of schizophrenia and of a particular attention response pattern. The observations of Singer and Wynne (1963) and Arieti (1955) concerning interpersonal relationships and communication between parents and their schizophrenic children are clearly consistent with the formulation presented here. While further systematic research is needed, such observations do appear especially noteworthy. Additional indications of the importance of early social environment on attention responsiveness are found in the process-reactive literature which is saturated with developmental inferences and analyses (e.g., Becker, 1956, 1959; Fine & Zimet, 1959; Rodnick & Garmezy, 1957). The field-differentiation hypotheses of Werner (1948) and Witkin et al. (1954, 1962) are most significant in this area (e.g., Bryant, 1961). Lastly, Piaget's (1950) conception of the developmental aspects of centration (scanning) has been found to be remarkably consistent with the process-reactive concept, at least in the early stages of schizophrenia (Harris, 1957; Zahn, 1959). However, the apparent inapplicability of the centration hypothesis to the attentional responses of long-term chronic process and reactive patients is, at first glance, disconcerting, to say the least. Clearly, the generality of the attention-response formulation is limited unless an explanation can be found for those studies in which chronic reactive patients evidence minimal scanning responsiveness as compared with chronic process patients. Even more intriguing is the question of why the scanning behaviors of chronic process patients are apparently no different from those of nonschizophrenic individuals (Zahn, 1959). These findings necessitate a consideration of the second of the two developmental issues, namely what might be the differential effects of prolonged hospitalization on patients having different premorbid social histories and perhaps different symptom patterns.

In order to deal with this latter problem, it will be necessary to postulate an attention mechanism which will be termed "sensory input processing–ideational gating." The data in support of this formulation come mainly from some neurophysiological findings by DeVault (1955) and from a study of reaction to perceptual ambiguity by Draguns (1963). The clinical observations of Bleuler on excessive registration of sensory impressions, of Arieti on awholism, and of Rapaport on hypermnesia are also especially well suited to interpretation in terms of this mechanism. The neurophysiological data indicate an autonomic activation pattern among process-chronic schizophrenics in which response to sensory stimuli is present but in which response to ideational stimuli and to noxious physical stimuli is strongly inhibited (negative change in heart rate, decreased GSR). Reactive-chronic schizophrenics show no autonomic inhibition to nonnoxious ideational stimuli. The perceptual ambiguity study indicates a preference among the more chronic schizophrenics to respond to the diffuse, undifferentiated aspects of stimuli rather than to their connotative attributes. On the basis of these findings, to be described in more detail below, the following formulation is presented:

1. Process-chronic schizophrenics characteristically minimize or "gate out" disturbing ideational inputs, both perceptual and conceptual, while at the same time responding to sensory inputs.
2. Reactive-chronic schizophrenics remain responsive to various ideational aspects of their environments as well as to sensory stimulation but they regress to a genetically earlier form of scanning responsiveness.[6]
3. Under extraordinary stress conditions all schizophrenics evidence indications of ideational gating.

DeVault (1955) studied the physiological response patterns (GSR, heart rate) of groups of chronic-process and chronic-reactive patients and of a nonschizophrenic group. Mean lengths of hospitalization for the two schizophrenic groups were 8.8 years and 8.7 years, respectively. Response patterns to a verbal warning followed (at least 30 seconds later) by a loud bell and to scenes of hostility, sex, dependency, and a "neutral" scene, were found to be significantly different for the three groups. The reactive-chronic group evidenced a significantly higher physiological arousal pattern than the process-chronic group. In contrast to the response pattern of the process group to the pictorial stimuli, the pattern of the reactive and nonschizophrenic groups corresponded to one normally observed when subjects are presented with ideational stimuli or noxious physical stimuli (Darrow, 1929; Jost, Ruilmann, Hill, & Gulo, 1952). The response pattern of the process group, however, resembled one not normally associated with reactions to ideational stimuli but rather one which is associated with responses to sensory stimuli (light, sound, etc.). This effect was noted in addition to the relative hyporeactivity of the process patients as compared to reactives and nonschizophrenics.

---

[6] Pearl (1962) has suggested that process schizophrenics, who he equates with chronic schizophrenics, are in a state of reduced stimulus input, whereas reactive schizophrenics, equated with acute schizophrenics, tend to be characterized by a state of stimulus overload. The general tenor of Pearl's hypothesis parallels the one presented here. However, note that the studies reviewed in this paper have indicated that the differentiation between process and reactive schizophrenia on the one hand and degree of chronicity on the other hand, which Pearl does not consider, is an extremely important distinction.

The hyporeactivity of the process group was not complete, however. The process schizophrenics did respond, but in a way different from that of the normals or reactives, with a deceleration (negative change) in heart rate after all the stimuli. This response, which according to Darrow is usually found after sensory stimuli, was often as large or larger than the positive changes in heart rate given by the normals and reactives. DeVault (1955) points out that during the presentation of warning and bell combination,

> where a negative change might have been expected and was found to occur in the normal group, (but not in the reactive group) the negative change of the process group was significantly larger than that of the normals, giving an indication of a stronger than normal response on this occasion [pp. 50–51].

Autonomic inhibition was also observed on the GSR polygraph recordings immediately after presentation of the verbal warning stimulus. Both schizophrenic groups evidenced a diminished GSR to the verbal warning when compared to the nonschizophrenic group. Since diminished GSR usually occurs in response to strong sensory stimulation rather than to strong ideational stimulation, this latter finding may indicate that all schizophrenics respond with "ideational gating" when they encounter stimuli with high "arousal power."

Draguns (1963) investigated the perceptual reactions of chronic and acute schizophrenics to ambiguous and sparse stimulus inputs. It was hypothesized that chronic patients as compared to acute schizophrenics and nonschizophrenics would demonstrate a strong disposition toward relatively premature and erroneous interpretations of perceptual data. That is, the chronic group was expected to respond to the diffuse, physiognomic qualities of the stimuli. Note that this prediction implied a receptive attitude to the sensory aspects of the data being presented by the experimenter.

One of the techniques employed to test this hypothesis utilized realistic drawings of objects or scenes which were photographically blurred to varying degrees. Nine series consisting of 12 pictures each were presented (beginning with the least clear presentation) to 34 chronic patients, 25 acutes in their first month of their first hospitalization, and 36 hospital attendants. (Subjects were equated for age, sex, amount of schooling, and verbal intelligence.) Each individual was instructed to respond at any point in a series, provided he "had an idea" of what the picture represented. Draguns (1963) found that the chronic group made their judgments at a significantly earlier point in time than did the nonschizophrenic group and that they made significantly more perceptual recognition errors as well. The performance of the acute group fell consistently between those of the chronic and nonschizophrenic groups.

These results are interpreted here as indicating that when arousal is increased experimentally due to the high uncertainty value of the stimulus data being presented (Berlyne, 1960), the schizophrenic will respond to the nonconnotative aspects of the inputs. Furthermore, this set will be taken despite having the knowledge, that adopting a more deliberate perceptual response orientation could, by itself, result in gradual uncertainty reduction. Such a "stylistic responsiveness to information input [Dragun's own term, p. 29]" is adopted in order (*a*) quickly to reduce uncertainty and (*b*) to render innocuous the potentially noxious qualitative aspects of inputs from the environment. To deal with experiences in this manner enables the schizophrenic to filter threatening connotations, i.e., to respond to reality on a piecemeal basis "so as to guarantee that no relationships will be met which will arouse impulses and anxieties" (Rapaport, 1951).[7] Sensory input processing–ideational gating is suggested as a mechanism through which such cognitive reorganizations take place in schizophrenia.[8]

SUMMARY

The data and theories reviewed indicate the following important relationships between the schizophrenic process and attention-response dispositions:

1. Extreme scanning and/or field-articulation perceptual dispositions characterize the attention response styles of most schizophrenics. Assessment of these response dispositions can be made by determining the constant error patterns of an individual's performances on a variety of psychophysical laboratory tasks.

2. Within the Kraepelinian categorization system, extensive scanning and field-articulation responsiveness are associated with the paranoid subtype, whereas minimal scanning and undifferentiated field-articulation responsiveness are associated with the catatonic, simple, and hebephrenic subtypes.

3. A moderate or nonextreme attention deployment response pattern in schizophrenia is associated with a good prognosis.

4. Within the process-reactive classification, undifferentiated field-articulation responsiveness is characteristic of process schizophrenics with poor premorbid developmental histories. Reactive schizophrenics evidence a more highly differentiated field-articulation responsiveness. These distinctions are consistent with the "genetic levels of cognitive development" hypothesis set forth by Werner, Piaget, and others.

5. The significance of the scanning response disposition in either of its extreme forms in schizophrenia derives on the one hand from its relationship to personality structure and on the other hand through its use in the analysis of basic perceptual orientations to the environment.

6. Long-term involvement in the schizophrenic resolution appears to be

[7] The concept of the "filtering" of sensory data has been discussed in detail by Broadbent (1958). It has also been applied to the re-evaluation of several studies of the effects of certain drugs on the narrowing and broadening of attention (Callaway & Band, 1958; Callaway & Dembo, 1958; Callaway & Stone, 1960). In fact, the modifiability of the scanning response via drugs suggests the possibility of an experimental chemotherapy program which could have important implications for the treatment of schizophrenia.

[8] The developmental and motivational bases for this mechanism are conceived to be: (*a*) initially, an acute psychological state of self-initiated sensory deprivation which McReynolds (1960) has postulated in his discussion of schizophrenic withdrawal, (*b*) a consequent reduction of sufficient sensory inputs in an organism who requires such inputs in order to exist, and (*c*) the gradual modification of a self-initiated sensory-deprivation "strategy" in terms of the processing of sensory input but the simultaneous nonprocessing of ideational input. The extent to which this mechanism is utilized is apparently dependent on premorbid history and clinical syndrome factors.

associated with change in attention response style. The kind of change is dependent upon the particular "style" which was operative during the premorbid and early phases of the schizophrenia.

7. The extensive utilization of an "ideational gating" mechanism which filters disturbing connotative environmental inputs is postulated as the ultimate schizophrenic defense mechanism.

8. The interactions between such important variables as the process-reactive distinction, physiological reactivity, degree of chronicity, anxiety levels, and Kraepelinian subtype are complex, to say the least. However, these interactions can be systematically explored within the attention response style model.

REFERENCES

Ardis, J. A., & Fraser, Elizabeth. Personality and perception: The constancy effect and introversion. *Brit. J. Psychol.,* 1957, **48,** 48–54.

Arieti, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.

Arieti, S. The loss of reality. *Psychoanal. Rev.,* 1961, **3,** 3–24.

Arieti, S. The microgeny of thought and perception. *Arch. gen. Psychiat.,* 1962, **6,** 454–468.

Becker, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.,* 1956, **47,** 489–496.

Becker, W. C. The process-reactive distinction: A key to the problem of schizophrenia? *J. nerv. ment. Dis.,* 1959, **129,** 442–449.

Berlyne, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

Bexton, W. H., Heron, W., & Scott, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.,* 1954, **8,** 70–76.

Bleuler, E. (Orig. publ. 1911.) *Dementia praecox or the group of schizophrenias.* (Trans. by J. ZinKin.) New York: International Universities Press, 1951.

Bockoven, J. S., Greenblatt, M., & Solomon, H. C. Social behavior and autonomic physiology in long-standing mental illness. *J. nerv. ment. Dis.,* 1953, **117,** 55–58.

Brenman, M., & Gill, M. M. *Hypnotherapy.* New York: International Universities Press, 1947.

Broadbent, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

Bryant, A. R. P. An investigation of process-reactive schizophrenia with relation to perception of visual space. Unpublished doctoral dissertation, University of Utah, 1961.

Callaway, E., & Band, R. I. Some psychopharmocological effects of atropine. *Arch. neurol. Psychiat.,* 1958, **79,** 91–102.

Callaway, E., & Dembo, D. Narrowed attention. *Arch. neurol. Psychiat.,* 1958, **79,** 74–90.

Callaway, E., & Stone, G. Re-evaluating focus of attention. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley, 1960.

Cameron, N. *The psychology of behavior disorders: A biosocial interpretation.* Boston: Houghton Mifflin, 1947.

Carlson, V. R. Underestimation in size constancy judgments. *Amer. J. Psychol.,* 1962, **75,** 462–465.

Crookes, T. G. Size constancy and literalness in Rorschach Test. *Brit. J. med. Psychol.,* 1957, **30,** 99–106.

Darrow, C. W. Differences in the physiological reactions to sensory and ideational stimuli. *Psychol. Bull.,* 1929, **26,** 185–201.

DeVault, S. H. Physiological responsiveness in reactive and process schizophrenia. Unpublished doctoral dissertation, Michigan State University, 1955.

Donahoe, J. W., Curtin, Mary E., & Lipton, L. Interference effects with schizophrenic subjects in the acquisition and retention of verbal material. *J. abnorm. soc. Psychol.,* 1961, **62,** 553–558.

Draguns, J. G. Responses to cognitive and perceptual ambiguity in chronic and acute schizophrenics. *J. abnorm. soc. Psychol.,* 1963, **66,** 24–30.

Escalona, S., & Heider, G. *Prediction and outcome.* New York: Basic Books, 1959.

Farrow, B. J., & Santos, J. F. Changes in autokinetic perception as a function of the transfer of conditioning effects. *Brit. J. Psychol.,* 1962, **53,** 331–337.

Fine, H. J., & Zimet, C. N. Process-reactive schizophrenia and genetic levels of perception. *J. abnorm. soc. Psychol,* 1959, **59,** 83–86.

Foulds, G. A., & Owens, A. Are paranoids schizophrenics? *Brit. J. Psychiat.,* 1963, **109,** 674–680.

Freud, S. Psychoanalytic notes on an autobiographical account of a case of paranoia (dementia paranoides). (Trans. by J. Strachey) Vol. 12. London: Hogarth Press, 1958.

Funkenstein, D. H., Greenblatt, M., & Solomon, H. C. A test which predicts the clinical effects of electric shock treatment in schizophrenic patients. *Amer. J. Psychiat.*, 1950, **106**, 889–901.

Gardner, R. W. Cognitive controls of attention deployment as determinants of visual illusions. *J. abnorm. soc. Psychol.*, 1961, **62**, 120–127.

Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, H. B., & Spence, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychol. Issues*, 1959, **1**, No. 4.

Gardner, R. W., Jackson, D. N., & Messick, S. J. Personality organization in cognitive controls and intellectual abilities. *Psychol. Issues*, 1960, **2**, No. 8.

Gardner, R. W., & Long, R. I. Cognitive controls of attention and inhibition: A study of individual consistencies. *Brit. J. Psychol.*, 1962, **53**, 381–388. (a)

Gardner, R. W., & Long, R. I. Control, defence and centration effect: A study of scanning behaviour. *Brit. J. Psychol.*, 1962, **53**, 129–140. (b)

Garmezy, N., & Rodnick, E. H. Premorbid adjustment and performance in schizophrenia: Implications for interpreting heterogeneity in schizophrenia. *J. nerv. ment. Dis.*, 1959, **129**, 450–460.

Grosz, H. J., & Miller, I. Two factors limiting the reliability of the mecholyl test: Variability in day-to-day responses and dissimilarity between simultaneous bilateral determinations. *J. nerv. ment. Dis.*, 1958, **127**, 417–429.

Guthrie, E. R. Association by contiguity. In S. Koch (Ed.), *Psychology: A study of a science.* New York: McGraw-Hill, 1959.

Hahn, I. N. The effects of success and failure experiences on chronic schizophrenic patients. Unpublished doctoral dissertation, University of Oklahoma, 1956.

Harris, J. G. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *J. Pers.*, 1957, **25**, 651–671.

Hartman, A. M. The apparent size of after-images in delusional and non-delusional schizophrenics. *Amer. J. Psychol.*, 1962, **75**, 587–595.

Havener, P. H., & Izard, C. E. Unrealistic self-enhancement in paranoid schizophrenics. *J. consult. Psychol.*, 1962, **26**, 65–68.

Held, R., & Hein, A. Adaptation of disarranged hand-eye coordination contingent upon reafferent stimulation. *Percept. mot. Skills*, 1958, **8**, 87–90.

Herron, W. G. The process-reactive classification of schizophrenia. *Psychol. Bull.*, 1962, **59**, 329–343.

Holway, A. H., & Boring, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **51**, 21–37.

Ilanit, N. Some psychological correlates of the process-reactive concept of schizophrenia. Unpublished doctoral dissertation, University of Southern California, 1961.

Jackson, D. *Etiology of schizophrenia.* New York: Basic Books, 1960.

Jost, H., Ruilmann, C. J., Hill, T. S., & Gulo, M. J. Studies in hypertension. I. Technics and control data. Central and autonomic nervous system reactions of normal adults to sensory and ideational (frustration) stimulation. *J. nerv. ment. Dis.*, 1952, **115**, 152–162.

Kantor, R. E., Wallner, J. M., & Winder, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157–162.

King, G. F. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *J. abnorm. soc. Psychol.*, 1958, **56**, 160–164.

Klein, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation: 1954.* Lincoln: Univer. Nebraska Press, 1954, 224–274.

Klein, G. S. Cognitive control and motivation. In G. Lindzey (Ed.), *Assessment of human motives.* New York: Rinehart, 1958. Pp. 87–118.

Kraepelin, E. *Dementia praecox and paraphrenia.* (Trans. by R. Mary Barclay.) Chicago: Chicago Medical Book, 1919.

Lovinger, E. Perceptual contact with reality in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 87–91.

Lynn, R. Russian research in schizophrenia. *Psychol. Bull.*, 1963, **60**, 486–498.

McConaghy, N. Modes of abstract thinking and psychosis. *Amer. J. Psychiat.*, 1960, **117**, 106–110.

McReynolds, P. Anxiety, perception, and schizophrenia. In D. Jackson (Ed.), *The etiology of schizophrenia.* New York: Basic Books, 1960, 248–292.

MEADOW, A., & FUNKENSTEIN, D. H. The relationship of abstract thinking to the autonomic nervous system in schizophrenia. In P. Hoch & J. Zubin (Eds.), *Relation of psychiatric tests to psychiatry.* New York: Grune & Stratton, 1952, 131-149.

MEADOW, A., GREENBLATT, M., FUNKENSTEIN, D. H., & SOLOMON, H. C. Relationships between capacity for abstraction in schizophrenia and physiological response to autonomic drugs. *J. nerv. ment. Dis.*, 1953, **118**, 332–338.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

PEARL, D. Stimulus input and overload in relation to classifications of schizophrenia. *Newsltr Res. Psychol.*, 1962, **4**, 44–56.

PIAGET, J. *The psychology of intelligence.* London: Routledge & Kegan Paul, 1950.

PIAGET, J. *The construction of reality in the child.* New York: Basic Books, 1954.

PIAGET, J., VINH-BANG, & MATALON, B. Note on the law of the temporal maximum of some optico-geometric illusions. *Amer. J. Psychol.*, 1958, 71, 277–282.

RAPAPORT, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.

RAPAPORT, D. On the psychoanalytic theory of motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1960.* Lincoln: Univer. Nebraska Press, 1960, 173–247.

RAUSH, H. L. Perceptual constancy in schizophrenia. *J. Pers.*, 1952, 21, 176–187.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957.* Lincoln: Univer. Nebraska Press, 1957, 109–184.

SANTOS, J. F., FARROW, B. J., & HAINES, J. R. How attention influences what is perceived: Some experimental evidence. *Bull. Menninger Clin.*, 1963, 27, 3–14.

SEXTON, M. C. The autokinetic test: Its value in psychiatric diagnosis and prognosis. *Amer. J. Psychiat.*, 1945, **102**, 399–402.

SHAKOW, D. Segmental set. *Arch. gen. Psychiat.*, 1962, **6**, 1–17.

SILVERMAN, J. Psychological deficit reduction in schizophrenia through response contingent noxious reinforcement. *Psychol. Rep.*, 1963, **13** (Monogr. Suppl. No. 2V13), 187–210. (a)

SILVERMAN, J. Cognitive controls and symptom patterns in schizophrenia. Unpublished research prospectus, Palo Alto, Calif.: Palo Alto Veterans Administration Hospital Library, 1963. (b)

SILVERMAN, J. The scanning control mechanism and "cognitive filtering" in paranoid and nonparanoid schizophrenia. *J. consult. Psychol.*, 1964a, in press.

SILVERMAN, J. Perceptual control of stimulus intensity in paranoid and nonparanoid schizophrenics. *J. nerv. ment. Dis.*, 1964b, in press.

SINGER, MARGARET T., & WYNNE, L. C. Differentiating characteristics of the parents of childhood schizophrenics, childhood neurotics and young adult schizophrenics. *Amer. J. Psychiat.*, 1963, **120**, 234–243.

SNYDER, S. Perceptual closure in acute paranoid schizophrenics. *Arch. gen. Psychiat.*, 1961, **5**, 406–410.

SOLLEY, C. M., & MURPHY, G. *The development of the perceptual world.* New York: Basic Books, 1960.

SOLLEY, C. M., & SANTOS, J. F. Perceptual learning with partial verbal reinforcement. *Percept. mot. Skills*, 1958, **8**, 183–193.

TAYLOR, J. N. A comparison of delusional and hallucinatory individuals using field dependency as a measure. Unpublished doctoral dissertation, Purdue University, 1956.

THOULESS, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, 22, 216–241.

VOTH, A. C. Individual differences in the autokinetic phenomenon. *J. exp. Psychol.*, 1941, **29**, 306–322.

VOTH, A. C. An experimental study of mental patients through the autokinetic phenomenon. *Amer. J. Psychiat.*, 1947, **103**, 793–805.

VOTH, H. M. Choice of illness. *Arch. gen. Psychiat.*, 1962, **6**, 149–156.

WAPNER, S., & WERNER, H. *Perceptual development: An investigation within the framework of sensory-tonic theory.* Worcester: Clark Univer. Press, 1957.

WEBER, C. O. The relation of personality trends to degrees of visual constancy correction for size and form. *J. appl. Psychol.*, 1939, 23, 703–708.

WECKOWICZ, T. E. Size constancy in schizophrenic patients. *J. ment. Sci.*, 1957, **103**, 475–486.

WECKOWICZ, T. E. Autonomic activity as measured by mecholyl test and size constancy in schizophrenic patients. *Psychosom. Med.*, 1958, **20**, 66–71.

WECKOWICZ, T. E., SOMMER, R., & HALL, R. Distance constancy in schizophrenic patients. *J. ment. Sci.*, 1958, **104**, 1174–1182.

WECKOWICZ, T. E., & BLEWETT, D. B. Size constancy and abstract thinking in schizophrenic patients. *J. ment. Sci.*, 1959, **105**, 909–934.

WEISS, R. L., & SHERMAN, M. Anxiety and interfering responses in college students and psychiatric patients. *Newsltr Res. Psychol.*, **4**, 35–40.

WERNER, H. *Comparative psychology of mental development.* Chicago: Follet, 1948.

WERNER, H., & WAPNER, S. The Innsbruck studies on distorted visual fields in relation to an organismic theory of perception. *Psychol. Rev.*, 1955, **62**, 130–138.

WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., KARP, S. A. *Psychological differentiation.* New York: Wiley, 1962.

WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

WYNNE, L. C., & SINGER, MARGARET T. Thought disorder and family relations of schizophrenics: II. A classification of forms of thinking. *Arch. gen. Psychiat.*, 1963, **9**, 199–206. (a)

WYNNE, L. C., & SINGER, MARGARET T. Thought disorder and family relations of schizophrenics: I. A research strategy. *Arch. gen. Psychiat.*, 1963, **9**, 191–198. (b)

ZAHN, T. P. Acquired and symbolic affective value as determinants of size estimation in schizophrenic and normal subjects. *J. abnorm. soc. Psychol.*, 1959, **58**, 39–47.

ZAHN, T. P. Size estimation of pictures associated with success and failure as a function of manifest anxiety. *J. abnorm. soc. Psychol.*, 1960, **61**, 457–462.

ZUCKERMAN, M., & GROSZ, H. J. Contradictory results using the mecholyl test to differentiate process and reactive schizophrenia. *J. abnorm. soc. Psychol.*, 1959, **59**, 145–146.

(Received April 26, 1963)

Psychological Review
1964, Vol. 71, No. 5, 380–391

A POSSIBLE OPTICAL BASIS FOR MONOCULAR SLANT PERCEPTION[1]

HOWARD R. FLOCK
Hunter College of the City University of New York

Optic variables that correspond to various properties of distal substances and define the axial coordinates and the slant of a surface are isolated in the light to the eye. The textural elements composing a surface are assumed to be randomly irregular in shape, size, and separation; and to have certain light-reflecting properties; but no prior assumptions are made about the position and illumination of the surface. The accurate specification of distal surface slant from optic variables is demonstrated both when the head of the viewer and the surface are motionless and when the surface is moving. Testable hypotheses that are derived from the theoretical model are mentioned at various places. Presumptive abilities necessary to the registration of the optic variables are also described.

Theoretically, accurate monocular slant perceptions are possible even though a motionless viewer has no previous experience with a particular substance, even though the textural elements of a motionless surface are irregular in size, shape, and separation, and even though parts of the surface are illuminated differently. This paper attempts to demonstrate the truth of this proposition, basing the demonstration on two hypothetical optic variables—the optical invariance of congruent luminous patterns, and optical *n*. The hypothetical basis for these two variables is shown to depend on two postulates about nature. To preclude covert introduction of unwanted assumptions, the abilities presumably needed to register the optic variables are outlined at each stage of the paper.

OPTICAL TEXTURE

Every point (station point) in and above terrestrial space is partly surrounded by surfaces. The light reflected from these surfaces travels in straight pathways and converges on individual station points as a solid spherical bundle of light rays. Gibson (1961) calls this bundle of light rays the *optic array*. Because of different incident lights falling on surfaces and because of the differential reflectances of substances, the light transmitted along different pathways in an optic array has different intensities and spectral values. The optic array therefore possesses a texture. This *optical texture* (see also Gibson, 1950a, 1950b, 1961) may be considered as a luminous pattern of relative intensities. It may be considered also as a single pattern or as subdivided into many individual patterns.

OPTICAL INVARIANCE OF CONGRUENT LUMINOUS PATTERNS

Consider the possibility (Postulate I) that substances (groups of substances, substantive structures, or substructures) of a particular class, possessing tangible regularities (e.g., congruencies and similarities) in layout and structure, have unique light-reflecting properties and reflect a

---

[1] This paper was supported in part by Public Health Service Research Grant, MH 06942-01, from the National Institute of Mental Health.

unique pattern of relative light intensities invariant to that class. As a corollary, substances of a different class reflect a different but equally unique pattern of relative light intensities.

Postulate I implies that the optical texture possesses spatially extended congruent luminous patterns that uniquely refer to substances of the same class. If it can be shown that the congruent luminous patterns tend to remain unchanged despite changes in the size, shape, position, and illumination of the distal substances, then Postulate I implies a concept of optical invariance.

Consider a collection of like substances—the grass blades in a lawn, for example. If two grass blades are identical in size and shape, are illuminated in exactly the same way, and are at the same distance from and slant to a station point, then both (according to Postulate I) should reflect congruent patterns of relative luminous intensities to the station point. Scanning one of these luminous patterns in steps of 30 seconds of arc, one might find the sequential spatial order of absolute luminous intensities to have the following (fictional) values:

60 15 20 35 60
15 20 35 60 . . . [1]

The pattern of relative changes in the light can then be described as:

4 1 1.3 2.3 4
1 1.3 2.3 4 . . . [2]

This luminous pattern of relative intensities refers to some part of the total luminous pattern reflected by a substance like a grass blade. If this luminous pattern of relative intensities is optically invariant, it must remain unchanged despite changes in size, shape, position, and illumination of the substance containing its distal referents.

Changing Incident Light. If the illumination on one blade is increased uniformly until the reflected absolute intensities at the station point are doubled (i.e., 120, 30, 40, 70, 120 . . .), the pattern of relative intensities remains invariant as in Sequence 2 above.

Changing Distance. When two grass blades are presented as before, but with one at a greater distance from the station point, the luminous patterns have the same relative intensities. The absolute intensities differ, of course, as a function of the inverse square law.

Magnification. Matter supposedly has almost infinitely fine structure. As one of the grass blades is brought closer to a station point, the angular sizes of its fine structure increase. In scanning part of a grass blade from a station point in steps of 30 seconds of arc, the absolute reflected intensities of light that are discriminable might be as in Sequence 1 above, and the relative intensities as in Sequence 2 above. At one half the distance, the absolute intensities for the same rays are: 240, 60, 80, 140, 240 . . . Because the angular separations of these light rays are approximately 60 seconds of arc, the new readings in steps of 30 seconds of arc might be: *240*, 160, *60*, 160, *80*, 160, *140*, 160, *240* . . . Correspondingly, at one fourth the distance the readings might be: *960*, 360, 640, 840, *240*, 360, 640, 840, *320*, 360, 640, 840, *560*, 360, 640, 840, *960* . . . The relative intensities are then: *4*, 1.5, 2.7, 3.5, *1*, 1.5, 2.7, 3.5, *1.3*, 1.5, 2.7, 3.5, *2.3*, 1.5, 2.7, 3.5, *4* . . .

At one fourth the distance a repetitive subcycle, 1.5, 2.7, 3.5, emerges in the optical texture. This subcycle could not be observed at the most extreme distance where the cycle 4, 1, 1.3, 2.3 alone was registered. At one fourth the distance, three classes of congruent luminous patterns can be identified, viz., 4, 1, 1.3, 2.3; 1.5, 2.7, 3.5; and the combination of these. Presumably, each of these classes of congruent luminous patterns refers to a unique level of substance. (It will be shown later that any one of these classes of luminous patterns can be used to specify the slant of the surfaces in which their distal referents are found.)

Changing Size and Shape. Changing the size or shape of a grass blade might affect some aspects of a luminous pattern, but the pattern of the subcycles should remain the same.

Gradations of Illumination. If illumination varies over each grass blade, luminous patterns for grass blades are not congruent. Nevertheless, small regions of a grass blade may remain under approximately constant illumination, although the level of illumination differs from region to region. In this event, congruent luminous patterns again can be observed in the optical texture: the congruent luminous patterns might correspond to the ribs of a grass blade (or of a group of grass blades), for example, instead of to a more molar level of structure.

Slanting a Surface. Two bits of identical matter may be illuminated in the same way and have the same light-reflecting properties, but because they lie in the slanted surface of some substance, they may be at different distances from a station point. In that event, the intensities of their reflected lights at the station point differ as a function of the inverse square law. These differences in absolute intensities are relatively trivial, however, for identical bits that lie within small regions on the surface.

Registration of the Optical Invariant. This section has set forth a concept of optical invariance for substances of the same class. When Postulate I holds true, the luminous patterns reflected by like substances tend to be congruent and unchanged despite changes in the position, size, shape, and illumination of the distal substances.

If this optical invariant is to be perceptually useful, visual systems must have the abilities necessary to register it. These include the ability to: (*a*) register relative changes and patterns of relative changes in luminous intensity; and (*b*) determine that one pattern of relative changes is the same as, or different from, a second pattern.

LINEAR REGULARITIES IN PHYSICAL SUBSTANCE

Congruent luminous patterns identify a distal collection of what are called hereafter, like elements. Like elements possess the same substantive structure but are not necessarily the same size and shape. In addition, adjacent like elements are not necessarily separated by the same distances. Although like elements may be irregular in size, shape, and separation, adjacent collections of like elements and their interspaces may be relatively uniform in size and shape. The truth of this proposition depends on a second assumption about nature (Postulate II). Although like elements (grass blades in a lawn, or the ribs of all grass blades in a lawn, etc.) vary in size, shape, and separation, such variations over a limited terrestrial region are normally distributed,[2] are random, and have a finite variance. This section shows how adjacent collections of like elements and their interspaces can be relatively uniform in size (see also Purdy, 1958).

---

[2] It is not clear that the assumption of normality is necessary: rectilinearity of the distribution may be sufficient.

Consider a lawn where the soil has relatively equal nutrients. Individual grass blades[3] probably differ in width and are separted by different distances. But, in accordance with Postulate II, these variations are random, are normally distributed, and have a finite variance. Measuring all of the widths and separations, one might find the combined mean width and separation of all grass blades to be .15-inch, with a standard deviation of .022-inch. This population of widths and separations could be sampled by stretching a wire in a straight line across a part of the lawn until it crossed, for example, nine grass blades and their separations. This procedure could be repeated at 100 other places on the lawn. Based on the population values, 95/100 of those linear extents would lie between approximately 1.22 and 1.48-inch, with a relative error of 9.6%. These limits are given by the following equation:

$$\text{Limits} = n(M_{\text{pop}} \pm 1.96\,\sigma/\sqrt{n}) \quad [3]$$

In this equation n is the numerosity of like elements in a collection; M_{pop} is the mean size and separation of the population of like elements; σ is the standard deviation of the population of sizes and separations; and 1.96 is a constant.

The relative error is given by the term: $(L_1 - L_2)/(L_1 + L_2)$, where L_1 and L_2 are the upper and lower limits, respectively. When the relative error is small, collections of n like elements are relatively regular in size. Relative error, therefore, is a measure of regularity and can be replaced by an equivalent expression that specifies the degree of regularity of a collection of like elements:

$$\text{Degree of regularity} = 1.0 - 1.96\,s/M\sqrt{n} \quad [4]$$

In Equation 4 s is the estimated standard deviation of the population, based on the sample variance; M is the mean of the sample; and n is the number of elements in the sample. From Equation 4 the degree of regularity depends on three variables and a constant. For example, if the sampling procedure described above is repeated for 36 instead of 9 widths and separations, the degree of regularity (using Equation 4) becomes .952 instead of .904. Increasing the sample size, therefore, increases the degree of regularity. The degree of regularity may be increased, also, by making s smaller (i.e., by making the elements more uniform in size, shape, and separation).

OPTICAL *n*

In Equation 4 the degree of regularity depends on three variables. Two of those variables, M and s, are indigenous to the particular matter, whereas the third variable, n, may be given any arbitrary value. What value n is given depends on the degree of regularity that is desired. The choice of a value for n is to be specified in accordance with optic criteria—the method of determination is described in the next section. But first an optic variable is described that corresponds to collections of n like elements.

Consider again the lawn described above. At a station point, the optical texture is packed with congruent luminous patterns, each corresponding to a grass blade. Select any luminous point (small luminous spot) in the optical texture, calling it the *optic fixation point*. With a line called the *optic line of regard*, connect the optic fixation point to the center of the optic array. Construct a meridian called the *optic meridian* through the optic fixation point. Along the meridian, count off a number, *n*, of contiguous congruent luminous patterns, calling this unit, *optical n*. Continue counting off such units until the meridian in the vicinity of the optic fixation point is subdivided into numerically equal units of *n*. These constructions are represented below in Fig. 1.

---

[3] Although I have based the analysis in this paper on grass blades that form a flat surface, the analysis could have been based on any collection of distinguishable substantive structures.

The numerosity of *n* in optical *n* corresponds exactly to the numerosity of *n* in the corresponding collection of *n* like elements. Whereas several collections of *n* like elements are relatively regular in size, however, the angular sizes of the optic units are almost never so. Only in those special instances when several collections of like elements are at the same distance from the station point are the optic units angularly uniform.

VALUE OF OPTICAL *n*

In order to specify accurately the slant of a distal surface from optic variables, optical *n* must be given a proper numerical value. That value is shown to depend on optic conditions which satisfy four criteria.

The optic meridian in the vicinity of the optic fixation point has been subdivided into numerically equal units of *n*. The angular sizes of the optic units may tend then to increase gradually along a segment of the meridian, to reach a maximum at the optic fixation point, and to decrease symmetrically in angular size. If these bilateral angular symmetries do not occur along the meridian, the meridian is to be rotated on the optic fixation point until they do occur. Whenever like elements lie in a plane and conform to Postulates I and II, bilateral symmetries may be observed along just one optic meridian through each optic fixation point.[4]

For some numerical value of optical *n*, the series of bilateral angular changes along the meridian may tend to reach a maximum at the optic fixation point, but the series may fail to have perfect symmetries. There may be reversals in the series. There may be bilateral asymmetries in the series, although there are no reversals. Nevertheless, the proper value of optical *n* is assumed to be that value of *n* that satisfies the following four criteria and correspondingly identifies relatively regular units of *n* like elements:

I. The series of gradually increasing, bilateral angular sizes along the optic meridian tends to reach a maximum at the optic fixation point.

II. The reversals and asymmetries in the series of angular sizes are distributed at random throughout the total series.

III. The numerical size of *n* is inversely proportional to the observed number of reversals and asymmetries in the total series.

IV. Other meridians parallel and close to the optic meridian yield the same value for *n* as does the optic meridian.

In those instances where Postulates I and II fail to describe nature accurately, a value of *n* may satisfy Criteria I to IV but the corresponding units of *n* like elements may not be

---

[4] There is an exception: when the optic line of regard, extended to intersect the surface, is normal to the surface, all optic meridians through that optic fixation point will yield these symmetries.

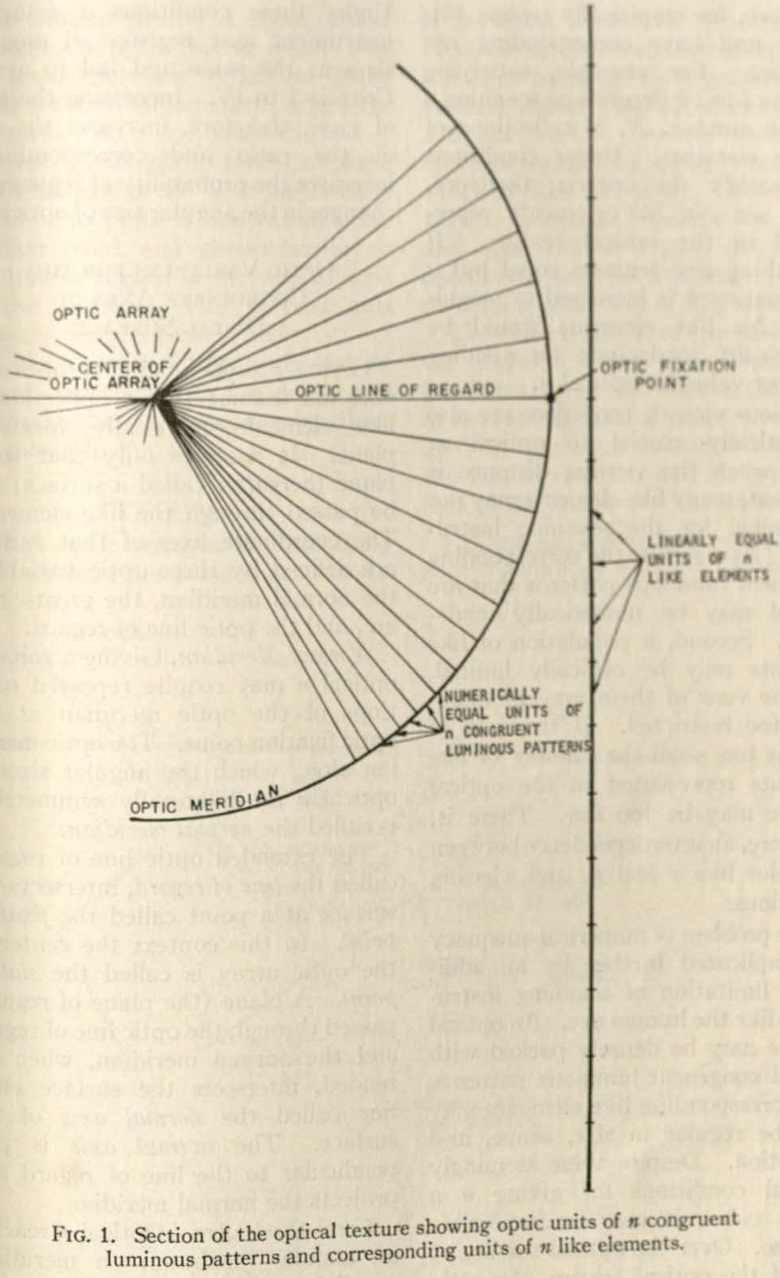

OPTIC ARRAY
CENTER OF OPTIC ARRAY
OPTIC LINE OF REGARD
OPTIC FIXATION POINT
LINEARLY EQUAL UNITS OF n LIKE ELEMENTS
NUMERICALLY EQUAL UNITS OF n CONGRUENT LUMINOUS PATTERNS
OPTIC MERIDIAN

FIG. 1. Section of the optical texture showing optic units of *n* congruent luminous patterns and corresponding units of *n* like elements.

regular. In those instances illusory perceptions might occur (e.g., Gibson & Flock, 1962). Any conditions that actually prevent or decrease the probability of satisfying the criteria should yield an improper value for *n* and, as a consequence, perceptual errors. Those conditions are a basis, therefore, for empirically testing this model and have corresponding importance. For example, satisfying Criteria I to IV depends on scanning a specific number, N, of collections of n like elements. Under conditions that satisfy the criteria, therefore, there are Nn like elements represented in the optical texture. If everything else remains equal but s in Equation 4 is increased in magnitude, Nn like elements would be numerically inadequate for yielding a proper value for optical n.

Various viewing conditions are also theoretically crucial to optical n. First, when the viewing distance is too great, many like elements may not be liminal for the scanning instrument. As a result, the corresponding congruent luminous patterns that are liminal may be numerically inadequate. Second, a population of like elements may be optically liminal, but the view of them may be angularly too restricted. If the field of view is too small the number of like elements represented in the optical texture may be too few. There is, therefore, an interdependence between variables like s and n, and viewing conditions.

The problem of numerical adequacy is complicated further by an additional limitation of scanning instruments like the human eye. An optical texture may be densely packed with liminal congruent luminous patterns. The corresponding like elements may even be regular in size, shape, and separation. Despite these seemingly optimal conditions for giving n a proper value, it may be impossible to do so. Over the optical texture or part of the optical texture the ratio of the largest to the smallest angular size of optical n may be greater than but close to 1.0. This can occur when an angular field of view is very small. Under these conditions a scanning instrument may register all angular sizes as the same and fail to attain Criteria I to IV. Increasing the field of view, therefore, increases the size of the ratio and correspondingly increases the probability of registering changes in the angular size of optical n.

OPTIC VARIABLES FOR THE COORDINATE AXES OF A DISTAL SURFACE

Giving a proper value to optical n does not assume that a collection of like elements physically forms a plane. It assumes only that some plane (hereafter called a surface) can be passed through the like elements. The coordinate axes of that surface are defined by three optic variables: the normal meridian, the great-circle arc, and the optic line of regard.

Normal Meridian. Giving a value to optical n may require repeated rotations of the optic meridian at the optic fixation point. The optic meridian along which the angular sizes of optical n are bilaterally symmetrical is called the *normal meridian.*

The extended optic line of regard, called the *line of regard,* intersects the surface at a point called the *fixation point.* In this context the center of the optic array is called the *station point.* A plane (the plane of regard) passed through the optic line of regard and the normal meridian, when extended, intersects the surface on a line called the *normal axis* of the surface. The *normal axis* is perpendicular to the line of regard and projects the normal meridian.

Great-Circle Arc. Optical n reaches an angular maximum on meridians parallel and close to the normal meridian (see Criterion IV above). A *great-circle arc* traced through these optic maxima is normal to the parallel meridians and passes through the optic fixation point. A plane passed through the optic line of regard and the great-circle arc, called the *plane of the slant*, when extended, intersects the surface on a line called the *line of rotation*. This plane forms a plane dihedral angle of 90° with the surface. Within this plane a line through the fixation point and perpendicular to the line of regard is the *vertical axis* of the surface. The vertical axis projects the great-circle arc.

When extended, the optic line of regard corresponds to the depth axis of the surface. These constructions are shown in Fig. 2.

OPTICAL *theta*

Each optic term in the preceding sections describes some characteristic of a distal collection of elements. First, the optical invariant (congruent luminous patterns) distinguishes substantive classes (like elements) in the material world. Second, optical *n* specifies regularities among randomly irregular like elements. Third, the normal meridian, great-circle arc, and optic line of regard define the coordinate axes of a surface passed through a collection of like elements.

The slant angle, θ, of the surface—the angle to which the surface has been rotated on its normal axis out of the frontal parallel plane—is measured by the angle included between the line of rotation and the vertical axis of the surface.[5] The optic vari-

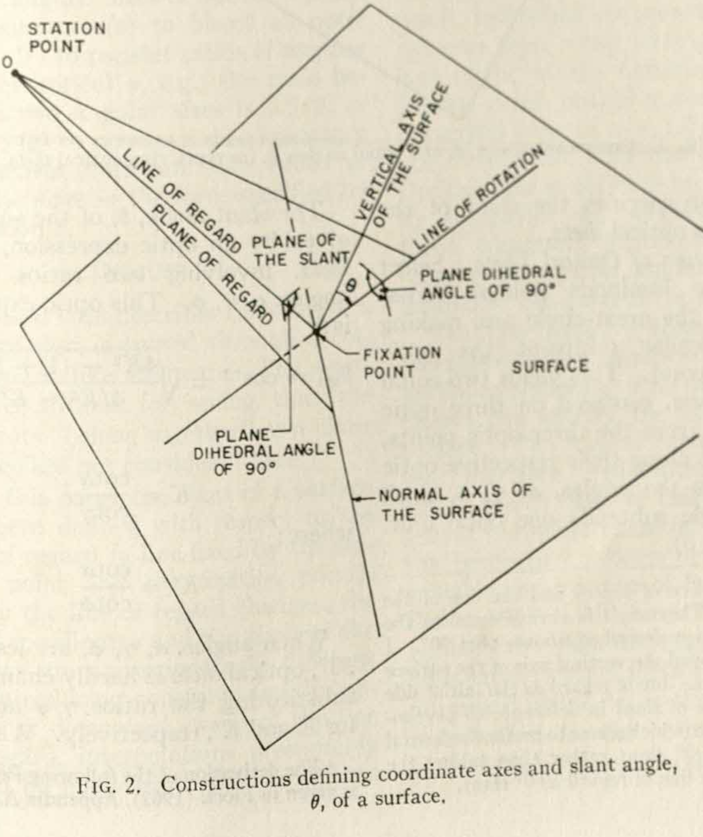

STATION POINT
O
LINE OF REGARD
PLANE OF REGARD
PLANE OF THE SLANT
VERTICAL AXIS OF THE SURFACE
LINE OF ROTATION
θ
PLANE DIHEDRAL ANGLE OF 90°
FIXATION POINT
SURFACE
PLANE DIHEDRAL ANGLE OF 90°
NORMAL AXIS OF THE SURFACE

FIG. 2. Constructions defining coordinate axes and slant angle, θ, of a surface.

---

[5] An alternate way to specify the angle of slant of the surface is to take the angle be-

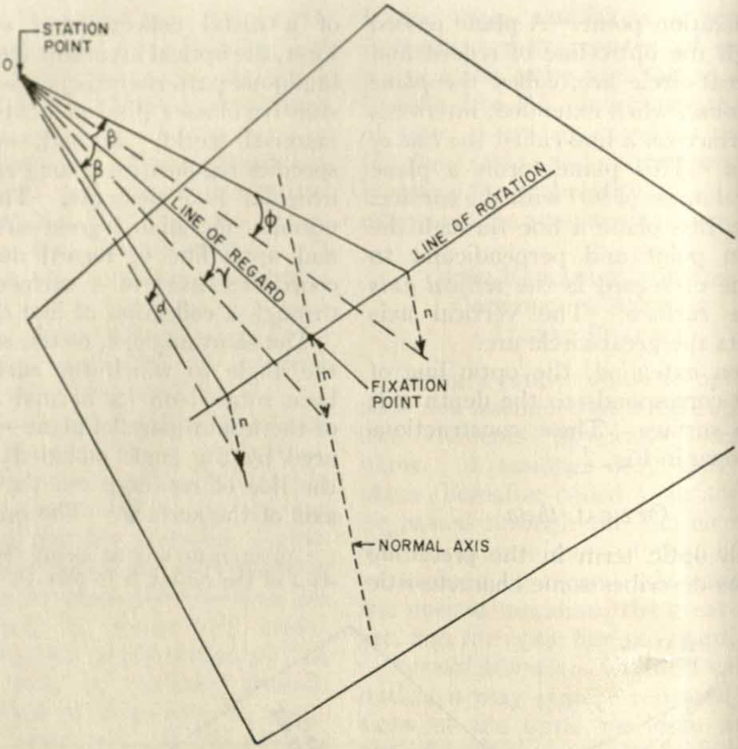


FIG. 3. Optic angles, α, γ, ϕ, and equal angles, β, for specifying optical *theta*.

able that specifies the slant of the surface is optical *theta*.

Derivation of Optical Theta. Select any two luminous points (spots) lying in the great-circle arc, making their angular midpoint the optic fixation point. This yields two equal angles, *beta*, centered on three optic points. From the three optic points, trace out along their respective optic meridians the angles, α, γ, ϕ, until each angle subtends one optic unit. (See Fig. 3).

The slant angle, θ, of the surface is given by an optic expression, *optical theta*, involving two ratios of the angles, α, γ, ϕ. This optic expression is:[6]

$$\theta_{\text{opt}} = \cos^{-1} \pm \sqrt{\frac{4K^2 - (1 + K')^2}{4(K^2 - K')}} \qquad [5]$$

$$K = \frac{\cot\alpha}{\cot\gamma}$$

where:

$$K' = \frac{\cot\alpha}{\cot\phi}$$

When angles, α, γ, ϕ, are less than 30°, optical *theta* is hardly changed by substituting the ratios γ/α and ϕ/α for K and K', respectively. When the

---

tween the line of regard and the line of the rotation. That angle is always equal to the angle of slant described above, plus 90°. I have employed the vertical axis of the surface instead of the line of regard as the initial side of the angle of slant in deference to psychological usage, which seems to prefer the frontal position as 0° slant rather than taking the plane of the line of regard as 0° slant.

[6] The derivation of the following expression is given in Flock (1962), Appendix A.

surface is flat, optical *theta* remains unchanged if the angles *beta* are taken increasingly larger and if angles α, γ, ϕ, subtend more than one optic unit. When the surface is curved, this process yields changing values for optical *theta*. Here, then, is a criterion for distinguishing flat from curved surfaces.

Registration of Optical Theta. In addition to the abilities presumed necessary for registering congruent luminous patterns, the additional abilities for registering optical n and optical *theta* are: (*a*) to trace out parallel meridians in the optical texture; (*b*) to count off equal numerical units of congruent luminous patterns along an optic meridian; (*c*) to observe relative differences in angular sizes of optical n; (*d*) to register optic loci where angular sizes of optical n reach a maximum; (*e*) to bisect an optic angle; (*f*) to register ratios of angular sizes of optical n, e.g., the ratio between two angular sizes is 3.5:1, or 3.7:1; (*g*) to possess a discriminatory mechanism that can take ratios of angular sizes in the form specified by Equation 5.

DISCUSSION

Optical *theta* describes the slant of a surface that is passed through a collection of like elements. The separate conditions for seeing that the elements "belong together" as a plane surface are not considered here.

In this paper the slant of a surface has been defined with respect to the line of regard (a line fixed by the station point and the fixation point). When the line of regard changes, the axial coordinates and the slant of the surface are correspondingly different By introducing constants for the upward or downward sweep of the line of regard, interpolations presumably could be made from slant as defined by the line of regard to slant as defined by gravity.

Optical n on which optical *theta* depends has been defined as a unit composed of several congruent luminous patterns and their luminous interspaces. Thus, a regularity among randomly irregular like elements was identified optically. The assumption that like elements are metrically irregular in size, shape, and separation may be more general than is sometimes required. Like elements (or their separations) sometimes may be metrically congruent along one or more dimensions but not along others. A surface might possess elements, for example, that are approximately the same size in a left-right direction, are randomly irregular in size in an up-down direction, and are distributed randomly over the surface. As a result, individual congruent luminous patterns lying along left-right meridians might satisfy Criteria I to IV. In that event, optical n would have a numerical value of $n = 1.0$ and would not include the luminous interspace. Optical *theta* could then be calculated by comparing the angular sizes of three congruent luminous patterns separated by equal angles *beta*, and located at the great-circle arc.

The light reflected from a surface includes different classes of congruent luminous patterns. Any one of these classes can specify the slant of the surface. The optical texture possesses, therefore, redundant information. Which class or classes of congruent luminous patterns is used by the perceptual process depends, presumably, on a range of factors (e.g., Flock, 1963). All things being equal perhaps the class of congruent luminous patterns which allows the smallest numerical value to optical n is the most salient for the perceptual process. At any rate, if perception involves a confirming and disconfirming scanning process, estimates of optical *theta* would be tested for different values of n as well as for different classes of congruent luminous patterns. Redundancy of information functions in this model, therefore, as a basis for confirmation, disconfirmation, and correction. Redundancy is not assumed to function in any additive or interactive way.

The formulation for optical *theta* assumes that both the viewer's head and the distal elements are motionless. The normal condition of organisms is to move; and the normal environment of an organism possesses movement. Nevertheless, a motionless viewer in a motionless world has more than academic interest. Beyond a few hundred feet there is little effective binocular disparity, and the human eyes are universally focused and not converged. Unless the organism is moving rapidly there is little effective motion parallax for distant objects. Except for rapidly moving "others," therefore, the optic conditions for the viewer looking at distant objects are essentially those of a motionless viewer in a motionless world. The motionless situation pertains, therefore, to distance perception and has corresponding practical importance.

If either the surface or the viewer is in motion the specification of optical *theta* is considerably simplified. It is to be noted that when a viewer moves with respect to some part of a surface, the optic conditions are exactly the same as when that part of the surface moves with respect to the viewer. The motion, whether it be of the surface or the viewer, therefore, can be referred to the surface.

The simplification of optical *theta* when there is motion can be described briefly. Consider a rigid surface that is in motion normal to the line of regard. Every point in the surface must be displaced through a linear distance, d, in time, t. If the surface is flat, the corresponding optic points have different angular velocities. All of the angular velocities in the optical texture travel along parallel optic meridians, reach a maximum at a locus of points called the great-circle arc, and bilaterally are at a minimum where the optic meridians converge. Select some optic point on the great-circle arc as the optic fixation point. Optical *theta* can be calculated from Equation 5 by comparing the angular displacements, α, γ, ϕ, for time, t, of three luminous spots which are separated by equal angles *beta* and are initially located on parallel meridians at the locus of the great-circle arc.

When the surface is in motion, Postulates I and II, congruent luminous patterns, and optical n do not enter the calculation of optical *theta*. Under these conditions the accuracy of estimation of optical *theta* presumably depends only on an optical texture in which the luminous changes are liminal, are sufficiently numerous, and are distributed over an extended optic area.

The theoretical model represented by optical *theta* yields numerous hypotheses that can be tested empirically. Some of those hypotheses have been and are being tested. Gibson and Flock's (1962) explanation of the mountain illusion, for example, is based on a systematic exception in nature to Postulate II. Flock and Moscatelli (1964) have found that increasing s in Equation 4 while holding constant the field of view and the numerosity of like elements yields less accurate slant judgments. They found also that once optical n has a minimal value, adding other regularities to a surface texture hardly improves slant judgments. Redundancy in the form of several classes of like elements also has been shown not to improve slant judgments (Flock, 1963). In an extended set of experiments with a surface in motion Flock (1964b) has shown that substituting good for bad textural shapes, changing velocity, prior exposure, concommitant perceptions of shape have little or no effect on the accuracy of slant judgments.[7]

The occurrence of slant perceptions is assumed to depend on optic variables. (For a more extended discussion of the assumptions underlying this point of view vis-à-vis, for example, Helmholtzian or Gestaltist orientations, see Flock, 1964a) One may know that a collection of distal elements does not conform to a slanted surface, and still one may see a slanted surface, as in the case of the cloud bank or of the virtual image on the cinema screen. The perceptual result depends on the saliency of the relevant optic variables. A perceiver may, of course, disbelieve his perceptions or ignore optic variables. Those are other problems and lead to other considerations.

At several places in this paper abilities are listed which are presumed to be necessary for the registration of the various optic variables. The level of difficulty implied by those abilities is, for the most part, that of making comparisons. How and when those abilities develop and how much experience a viewer must have with the optic variables before the abilities emerge are presumably empirical questions.

REFERENCES

Flock, H. R. *The monocular perception of surface slant.* (Doctoral dissertation, Cornell University, 1962) Ann Arbor: University Microfilms, 1962, No. 62-2514.

Flock, H. R. Selective registration of information in visual judgments of surface slant. *Percept. mot. Skills*, 1963, **17**, 537–538.

Flock, H. R. Three theoretical views of slant perception. *Psychol. Bull.*, 1964, **62**, 110–121 (a).

Flock, H. R. Some conditions sufficient for accurate monocular perceptions of moving surface slants. *J. exp. Psychol.*, 1964, **67**, 560–572. (b)

Flock, H. R., & Moscatelli, A. Variables of surface texture and accuracy of space perceptions. *Percept. mot. Skills*, 1964, **19**, 327–334.

Gibson, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950. (a)

Gibson, J. J. The perception of visual surfaces. *Amer. J. Psychol.*, 1950, **63**, 367–384. (b)

Gibson, J. J. Ecological optics. *Vis. Res.*, 1961, **1**, 253–262.

Gibson, J. J., & Flock, H. R. The apparent distance of mountains. *Amer. J. Psychol.*, 1962, **75**, 501–503.

Purdy, W. C. *The hypothesis of psychophysical correspondence in space perception.* (Doctoral dissertation, Cornell University, 1958) Ann Arbor: University Microfilms, 1958, No. 58-5594.

(Received June 11, 1963)

---

[7] Many more of the hypotheses mentioned throughout this paper have been tested by the writer and the data are being prepared for publication.

Psychological Review
1964, Vol. 71, No. 5, 392–407

CONSISTENCY OF AUDITORY DETECTION JUDGMENTS[1]

DAVID M. GREEN

Psychology Section and Research Laboratory of Electronics,
Massachusetts Institute of Technology[2]

An O's decision in a psychoacoustic detection experiment is governed by 2 broad classes of determinants: (a) external determinants, such as the likelihood of a particular wave form being a signal in noise and (b) internal determinants, such as the momentary state of the O's nervous system, his response biases, or biases as to certain sequences of responses. An analysis of these determinants and especially their relation to a trial-by-trial prediction of the S's response (a molecular psychophysics) was undertaken. The analysis suggests that the relative importance of the external and internal determinants can be assessed by measuring the consistency of Os' responses to repeated presentations of the same auditory signals. An experimental technique is introduced to measure such consistency and the results of some experiments are summarized. The results indicate that the amount of external variability is about the same as the internal variability. The implication of this result for any attempt to predict trial-by-trial behavior is discussed.

The ultimate goal of psychoacoustics is to understand the relation between the acoustic stimulus and the observer's reaction or response to this stimulus. Unfortunately, this relationship can be influenced by a variety of nonacoustic variables. Even in as simple a situation as a detection experiment in which the subject is asked whether or not a certain signal is present, the observer's response is determined by his criterion for the presence of the signal. This criterion, in turn, is determined by motivational and instructional variables that are, strictly speaking, nonstimulus variables.

Although detection theory has contributed a means of specifying the criterion, the increase in the amount of data needed to achieve this specification has almost precluded its adoption as a standard technique in sensory psychology. In an effort to bypass the criterion problem in its entirety, some experimenters advocate that a forced-choice method be used. In a two-alternative, forced-choice test, for example, the subject is presented with two temporal intervals and asked to state in which of the two intervals a signal was presented. Any asymmetry in the subject's choice of the two intervals would then be explained, presumably on the basis of some asymmetry in the sensory processes or some response biases. The problem here is that, although response biases or response preferences have long been studied, there is still ambiguity about the magnitude of these effects in typical sensory experiments. The ambiguity exists largely because most of the evidence

---

[1] This work was supported in part by the United States Army, the Air Force Office of Scientific Research, and the Office of Naval Research; and from the Air Force (Operational Applications Laboratory) under contract AR 19(604)-7459; and in part by the National Science Foundation (Grant G-21807).

[2] Now at the Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania.

for the operation of these response mechanisms occurs in experimental situations where relevant sensory information is held at minimal levels.

The immediate goal of this paper is to assess the relative contribution of these nonstimulus variables in a typical psychoacoustic detection task, one in which the auditory stimuli, at least on some of the trials, are exercising considerable influence on the observer's behavior. If the extent of such nonstimulus variables could be measured, then one might try to devise experimental techniques aimed at minimizing the influence of such factors in psychophysical experiments. Alternatively, it may become apparent that response biases are an integral part of all psychophysical phenomena and that a recognition of these biases and their operation is essential for meaningful interpretations of psychophysical results. In the latter case this technique provides a first step in the consideration of such factors. Finally, as a more remote goal, is the hope that the present investigation will provide some information on the feasibility of developing a more molecular approach to psychophysical phenomena. This last aim will become apparent later in the paper.

The paper begins with an analysis of the nonstimulus variables. That, in turn, leads to the introduction of a particular experimental technique, which, it is then argued, provides a realistic estimate of the influences of these nonstimulus variables in a psychophysical experiment. Preliminary results obtained with this technique establish an upper bound on the influence of stimulus variables in a two-alternative, forced-choice detection task. Unfortunately, the exact character of the nonstimulus variables is still ambiguous. We cannot, for example, at present distinguish between pure response preferences and internal or biological noise. A further extension of the technique is proposed that should clarify this ambiguity.

Before we consider the nature of the nonstimulus variables, let us review briefly the typical subject matter of psychophysical research.

MOLAR PSYCHOPHYSICS

Consider a typical psychophysical experiment and the procedures used in collecting and presenting data. One often obtains large numbers of responses to various stimulus intensities until some critical value of stimulus intensity (for example, the threshold intensity) is determined. Once the estimate of this intensity is obtained, the experiment is continued by changing some aspects of the situation, such as the frequency or duration of the signal, and repeating the basic procedure to obtain another estimate of this critical intensity. Finally, a graph is plotted with this critical intensity of the signal as the dependent variable and the other physical parameter as the independent variable. Thus, for example, the results show how the intensity of the signal must be changed as a function of signal duration to preserve the just detectable character of the signal. Note that the relation is between two physical aspects of the stimulus; the parameter is the average of a set of responses. Clearly, the inferred results apply only to average responses and not to the responses of the individual subject on a trial-by-trial basis. Thus most of traditional psychophysics is what we shall call "molar" psychophysics, to emphasize the point that the basic subject matter is an average of many individual psychophysical judgments.

It is often convenient to forget this fact and, indeed, many of the models in sensory psychology might be taken to be models that would or could be used to predict individual responses. In fact, no such models have ever been used in this way[3] and almost all of the basic data of psychophysics pertain to averages or collections of responses rather than to individual responses.

The necessity for this molar, rather than molecular, analysis is obviously dictated by our tacit understanding that the individual responses may at times be influenced by variables other than the physical characteristics of the stimulus. Two broad categories of these nonstimulus variables can be identified. One category is internal noise, the presence of random fluctuations within sensory systems that are claimed to influence the observer's judgments. The other broad category of nonstimulus variables encompasses a variety of supposed mechanisms, such as pure response mechanisms, time-order effects, stimulus expectancies, and the like. The communality of these diverse mechanisms is that all of their effects are derived from an analysis of the sequence of past responses. Thus they agree theoretically that an analysis of the past sequence of responses can be used, to predict the next response.

Molar psychophysics, by averaging over many responses, is used to remove or suppress these nonstimulus influences, and hence to reveal the "pure" stimulus determinants of the response. Frequently several subjects are employed, and the results from the different subjects averaged to suppress effects particular to individual subjects; then the analysis is not only molar but, strictly speaking, fails to apply to any individual subject. Rather, such results apply to some collection of subjects or some average observer.

MOLECULAR PSYCHOPHYSICS

Given the preceding designation of molar psychophysics, the subject matter of molecular psychophysics must be the individual response of the individual subject. The virtues of such a level of analysis are many, but two deserve special comment. First, a trial-by-trial analysis of the observer's response may provide a critical and efficient test of a hypothesis that is impossible on the molar level—impossible either because there is no difference in prediction on that level or because the amount of data required to select one hypothesis over another would be excessive. One can think of examples in which the relative efficiency of the molecular level of analysis is greater by the number of responses used to determine the average response on the molar level. Second, and more pertinent to this paper, molecular analysis provides a rigorous and demanding test of ideas about the manner in which nonstimulus variables influence the judgment process. Understanding one of these variables on the molecular level is equivalent to being able to predict its influence on individual responses. Understanding the operation of a nonstimulus variable on the molar level is equivalent to being able to reduce some asymmetry in the subject's average judgments by elaborate counterbalancing procedures. If the initial asymmetry is not very great, the effectiveness of these elaborate experimental procedures is frequently difficult to determine.

The ultimate success of molecular psychophysics depends, of course, on how large a portion of the individual response is determined by the stimulus, or by easily monitored response processes. If the individual responses are largely determined by internal noise, then a prediction of trial-by-trial responses will never be more than marginally successful. It is possible, for example, that Sherwin, Kodman, Kovaly, Prothe, and Melrose (1956), who attempted to predict individual response, were largely unsuccessful because there was too much internal noise. Alternately, their simulation of the sensory process might have been wrong, or perhaps their weakness was incorporating only stimulus variables in their prediction scheme. These are some of the alternatives that this paper tries to assess. Let us return to our two broad categories of nonstimulus variables and indicate the manner in which they relate to the prediction of individual responses.

---

[3] Sherwin, Kodman, Kovaly, Prothe, and Melrose (1956), is the singular exception to this generalization. Unfortunately, their attempt to predict trial-by-trial behavior was not very successful. The present paper was heavily influenced by their work and, in fact, might be regarded as an analysis of the reasons for their lack of success. For earlier work on the subject, see Aron and Irwin (1932), Preston (1936a), Preston (1936b).

SEQUENCE OF PAST RESPONSES

There is little question that the subject is influenced by the sequence of his past responses. What happened on the past response, or the conjunction of what happened on the past response and whether or not that response was correct, certainly will have *some* effect upon the response given on the present trial. There has been ample experimental demonstration of the statistical significance of such factors.[4] The question at issue is the size of these effects; what is the relative importance of these variables in actual psychophysical tests?

If response mechanisms could be understood in detail, then the experimenter, simply by tabulating the past responses, could work out the tendency for various responses and, by combining this variable with the stimulus factors, could, in theory, predict the trial-by-trial behavior of the subject. Actually, of course, there are very few available models that even attempt to deal with both stimulus and response variables. Those theories that achieve the most definite expression of the response factors are usually vague in their characterization of the stimulus and vice versa.[5]

There is, in addition to this lack of theoretical structure, a rather difficult experimental problem that has never been fully faced. The problem is that practically all of the data demonstrating the importance of response factors has been collected in situations in which there is little or no sensory stimulus. The experimental situation is often equivalent to asking the subject to guess the location of the stimulus in a two-alternative, forced-choice task with the stimulus intensity set to zero. In such a situation, the observer has little chance to utilize any sensory information, since it is held to a minimal level throughout the experiment. Consequently, it is not too surprising that the subject's responses are statistically dependent.[6] Unless the subject can generate completely random sequences, no other outcome would be expected. Furthermore, setting the signal to zero may produce a bias in and of itself. If there is no signal, then there is no means of reinforcing the subject for making subtle stimulus distinctions. There is, in fact, no way of insuring that the responses obtained from the subject are even partially determined by the same class of variables that operate when the subject is making stimulus distinctions. The subject may, as the test continues, begin to generate intricate sequences of responses simply to satisfy some whim or fancy of his own. No objective indication of this change in the observer's behavior will occur. Indeed, from the experimenter's point of view, the emergence of certain regularities in the experimental data may be very reinforcing, and it does not seem wholly impossible that this pleasant state of affairs is communicated to the observer in subtle ways. Even dismissing this possibility, turning the stimulus off in a psychophysical experiment seems to satisfy all of the formal requirements for the generation of superstitious behavior, as it has been studied in lower animals (Skinner, 1948). Superstitious behavior is, after all, nothing more than an elaborate stereotyped repetition of certain sequences of responses. Thus in a very real sense the data obtained in psychophysical experiments without a sometimes detectable signal may have very little predictive value in typical psychophysical experiment.[7]

---

[4] Senders and Soward (1952) contains a good review of the earlier published works. McGill (1957) summarizes much of the most recently published data. Apparently the first person to directly investigate the independence of judgments within the method of constant stimuli was Fernberger. For earlier work on subject, see Aron and Irwin (1932), Preston (1936a), and Preston (1936b).

[5] See Luce (1959). The stimulus is summarized by a multiplicative constant that, when greater than unity, increases one response tendency relative to another. Whether the constant refers to signal power, logarithm of signal power or some other transformation of the signal wave form is left as a problem to be determined empirically. This vagueness about the stimulus in response theories is counterbalanced by sensory or stimulus theories that, since there is no absolute need to represent response biases within the theory, usually ignore them.

[6] Responses obtained in ESP experiments have often been analyzed to determine re-

On the other hand, if the subject is achieving nearly 100% correct responses in a two-alternative, forced-choice task, then the stimulus influence is obviously so strong that the response factors cannot manifest themselves in the data. Whether the manifest and latent levels go hand in hand is largely an unanswered question. One might maintain, for example, that as the stimulus factors become more influential, the response factors are driven to zero; or one can defend the position that as the stimulus factors become more influential, the presence of response factors become harder to detect in the data, but that on the latent level they are as powerful as ever. Each position has some testable consequences, but little work has been carried out on this problem.

As previously observed, all of these determinants, simple response biases, time-order effects, motivational changes brought about by making a correct response on the previous trial, and so forth, while different on a theoretical level, are measured and inferred from the history of previous responses. They are operationally similar and thus, for the purposes of this paper, we shall combine all of these effects under the broad category of response-sequence effects.[8]

---

sponse sequence characteristics. See Goodfellow (1938), Skinner (1942), and Solomon (1949). The paper by Speeth and Mathews (1961) illustrates the use of a very low intensity signal in studying response factors.

[7] The second conclusion of the work of Senders and Sowards (1952) is as follows: "no subjects should be required to make threshold or near threshold discriminations for many times in succession. Under these conditions only extremely patient, well-trained subjects can continue to let their responses be determined by the stimulus alone; others will develop a pattern of responses and each response will tend to be response-determined rather than stimulus-determined. If threshold measurements are to be made, easier discriminations should be interspersed [p. 373]."

[8] This designation of response sequence effects is very nearly what McGill (1957) called "guessing." "'Guessing' then becomes a label for effects attributable to a broad

INTERNAL NOISE

Finally, we come to a concept that has had a very long history in psychophysics. Historically, internal noise, some random perturbation of the sensory processes, was the main impetus toward the development of psychophysical methods. Internal noise is an inferred quantity, its presence deduced from its manifest effects. It cannot be observed directly even by physiologists, for until the sensory code is completely understood, one can hardly distinguish what part is noise and what part signal. Thus, presently, on both physiological and psychological levels, noise is synonymous with the presence of effects that are not fully understood.

On an operational level, internal noise is equivalent to the observation that the same physical stimulus may elicit different responses. In a sense, then, internal noise is the limiting factor in a trial-by-trial prediction of the subject's response.

It should be noted that response factors and internal noise are similar in that both prevent a complete prediction of the subject's response from stimulus factors alone. They differ largely in that presumably, a more complete understanding of the response-sequence mechanisms and a successful combination of these response tendencies with the stimulus variables would allow one to predict the trial-by-trial behavior of the subject. Internal noise effectively prevents such a prediction, since there is no way, by definition, to monitor the effects of this variable.

AIM OF THE PRESENT RESEARCH

The aim of this paper is to develop an experimental technique that will allow one to estimate the magnitude of the nonstimulus factors and to assess, at least crudely, the relative importance of stimulus and nonstimulus factors. The present experimental results do not completely achieve the aim. The experimental technique does, however, have the *potential* to apportion the relative variability caused by internal noise and response dependencies, and we shall return to this potential later in the paper. The basic rationale of the procedure can be quickly summarized. If the stimulus alone determined the subject's response, and the subject were presented with the same stimulus on two separate occasions, we would expect the same response. Even if each individual judgment were subjected to some variability, we might expect the judgments to *tend* to be in agreement. Obviously, the exact amount of agreement, the degree of consistency, would depend on the relative influence of the stimulus factors and the amount of this internal variability. Let us now consider the procedure.

EXPERIMENTAL PROCEDURE

A two-alternative, forced-choice procedure was used throughout the experiments. Two types of trials were presented to the observer. One type, a signal trial, had a signal added to the noise background. The signal occurred at random in one of the two time intervals of the trial. After the subject made his response, he was informed whether or not it was correct. In the other type of trial, a no-signal trial, no signal was added in either of

---

class of determiners including set, previous training, method of responding, and so on. The essential feature of this class of constraints is that its effects in judgments are the same no matter what stimulus has been presented to the subject. In short, guessing habits act independently of stimuli. It follows that guessing can be measured by measuring the predictability of successive judgments . . . [p. 297]."

the two temporal intervals of the forced-choice trial. Thus the observer simply listened to two samples of noise. No information followed a no-signal trial, since no "right" or "wrong" response is possible.

A sequence of 100 trials constituted a single experimental block, which lasted approximately 5 minutes and consisted of 50 signal and 50 no-signal trials.[9] The purpose of mixing signal and no-signal trials was (*a*) to provide the observer with a constant reminder of the type of signal used in the psychophysical test and a means of reinforcing appropriate responses to that signal and (*b*) to monitor the level of the observer's performance throughout the test sessions.

The dependent variable of the experiment is based on a comparison of the subject's response during the first and second presentations of the tape. Consider only the no-signal trials. If the subject selects the same interval as his response on both the first and second presentations of the audio tape recording, then this constitutes an agreement; if the responses differ, there is no agreement. The percentage of agreements is calculated from the ratio of agreement trials to the total number of no-signal trials. This percentage will be the main dependent variable of the experiments discussed here.

All audio information was recorded on a single channel of the magnetic tape; on the second channel the timing information, that controls the various visual indicators that accompany the auditory information, was stored. This dual channel arrangement ensures almost perfect synchrony between the occurrence of an audio event and a corresponding observation-interval light. This correspondence is essential since the signal duration was only 1/10 second.

An Ampex Model 320 tape recorder was used. The audio information on the tape had a signal-to-noise ratio in excess of 40 decibels. While it is impossible to guarantee that the two presentations of the tape were exactly alike, every precaution was taken to approach that goal. The largest source of stimulus variability probably arose because of variation in the speed of the tape through the heads and slight variation in contact between the playback head and the magnetic tape. These variations produce both amplitude and frequency modulation of the audio signal.[10]

EXPERIMENTAL RESULTS

Two main experiments were conducted to probe the manner in which differences in the details of the experimental procedure might influence our estimate of the observer's consistency. The two experimental questions were: (*a*) How does the level of the signal presented on the signal trial influence the consistency measured on the no-signal trial? (*b*) How does the character of the signal and the exact presentation procedure employed on the no-signal trials influence the observer's consistency?

Experiment I

Noise was used as the signal in this experiment, that is, the signal was an increment in power of the continuous background noise. Three levels of increment for the signal were selected so that the subjects would receive approximately 60, 75, or 90% correct at the highest signal level. Four tapes, each containing 100 trials, were prepared at each signal level. The subject listened to each tape six times. Percentage agreement scores were calculated by comparing the response given on successive presentations of the tape. The six presentations yielded three independent estimates of a per cent agreement score.

---

[9] Presentation order was determined by random sampling without replacement, from a set of 100 items—half even, half odd.

[10] Certain aspects of the taped signal (power or energy level in the signal interval) were repeated almost exactly since for certain recordings simple electronic detection systems (broad-band energy detectors) were able to achieve 100% consistency in judgments based on two presentations of the tapes.

A two-stage sampling process is involved in this procedure. By repeating the same recording several times and determining the percentage agreement for successive pairs of presentations, we obtain an estimate of the percentage agreement score for that particular tape sample and its particular pairs of noise stimuli. By using several different tapes, we obtain different estimates from the population of stimulus pairs and hence estimates of the consistency that is possible in this particular physical situation.

In the present instance, our estimate of the percentage agreement for each *tape* is based on 150 trials (50 no-

TABLE 1

PERCENTAGE AGREEMENT SCORES AND PERCENTAGE CORRECT DETECTION SCORES (IN PARENTHESES) BASED ON 150 OBSERVATIONS

	Weak signal			
	Observer 1	Observer 2	Observer 3	Mean over observers
Tape 1	69 (57)	57 (55)	69 (57)	65 (56)
2	71 (68)	63 (62)	53 (61)	62 (64)
3	66 (58)	66 (55)	67 (63)	66 (59)
4	69 (61)	70 (60)	65 (57)	68 (59)
Mean over tapes	69 (61)	64 (58)	64 (60)	65 (60)

	Medium signal			
	Observer 1	Observer 2	Observer 3	Mean over observers
Tape 1	61 (82)	64 (77)	54 (76)	60 (78)
2	73 (69)	60 (76)	58 (75)	64 (73)
3	75 (68)	71 (77)	63 (68)	70 (71)
4	68 (73)	65 (73)	66 (68)	66 (71)
Mean over tapes	69 (73)	65 (76)	60 (72)	65 (74)

	Strong signal			
	Observer 1	Observer 2	Observer 3	Mean over observers
Tape 1	61 (93)	60 (91)	52 (88)	58 (91)
2	69 (94)	62 (92)	62 (93)	64 (93)
4[a]	69 (90)	59 (89)	66 (91)	65 (90)
Mean over tapes	66 (92)	60 (91)	60 (91)	62 (91)

[a] Tape 3 was excluded because detailed analysis indicated that slight amplitude distortion of the noise was present. The obtained scores were Observer I, 57(84); Observer II, 60(82); Observer III, 56(82).

signal trials, and 3 pairs of observations). Our consistency estimate for this physical situation is based on 200 sample pairs of noise stimuli (50 no-signal trials per tape, and four tapes).

The results of this experiment are given in Table 1 in which the average percentage of correct detections for all six presentations of the tape is presented together with the average percentage agreement score based on three pairs of presentations of the tape.

The grand means for the three conditions are indicative of the main conclusion, namely, there does not seem to be a strong relation between the percentage of correct detections obtained on signal trials and the consistency of the observer's judgments as inferred from the no-signal trials. Each individual observer supports this conclusion.

We shall postpone a more detailed discussion of how response dependencies might influence the percentage agreement score until the remaining experimental results have been discussed.

Experiment II

Consistency with Sinusoidal Signals

The main question explored in this series of experiments was: Does the consistency of the observer's judgments depend heavily on the kind of signal used in the tests? In the results just discussed, the signal was an increment in the level of a noise. In this experiment the signal was a gated sinusoid of 1/10-second duration. Four signal frequencies were investigated, 250 cps, 500 cps, 1,000 cps, and 2,000 cps. At each frequency four tapes were prepared and the subject listened to each tape four times. Table 2 lists the results obtained in these experiments.

The average percentage agreement is approximately 70% at all frequencies except 1,000 cps for which the average is 78%. All subjects had participated in another experiment of several weeks duration in which the signal frequency was 1,000 cps, and this may account for the slightly higher percentage agreement score obtained at this frequency. Note, however, that this result is almost wholly determined by the results of the third subject.

Other Signal Conditions and Presentation Procedures

In addition to the two stimulus conditions described thus far, a variety of other presentation procedures were investigated. Most of these situations were not tested exhaustively, although an attempt was made to sample the variety of different procedures in order to determine if any one would produce outstandingly consistent behavior.

The following conditions were investigated:

1. Tapes with silence between the observation intervals. The pulse audio wave form of 1/10-second duration was presented during the observation interval. The wave form was either noise alone or signal plus noise.

2. A pedestal was employed. In both observation intervals of the forced-choice procedure, a stimulus of the same kind as the signal was added to the background noise. The signal was then added in phase to this constant increment in one of the two intervals on the signal trial. Thus the subject's task was to select the more intense of the two pulsed sinusoids.

3. The same signal was presented in both intervals of the forced-choice procedure. Instead of 50 trials with no signal in either interval of the forced-choice procedure, a tape was prepared with 50 trials in which the same signal

TABLE 2

PERCENTAGE AGREEMENT SCORES BASED ON 100 OBSERVATIONS

	Observer 1	Observer 2	Observer 3	Mean over observers
Signal 250 cps				
Tape 1	68	72	70	70
2	80	72	68	73
3	76	54	78	69
4	76	52	78	69
Mean over tapes	75	62	74	71
Signal 500 cps				
Tape 1	84	67	45	65
2	90	82	51	74
3	74	64	57	65
4	70	78	70	73
Mean over tapes	80	73	56	69
Signal 1,000 cps				
Tape 1	76	77	85	79
2	86	73	79	79
3	83	75	77	78
4	76	71	79	75
Mean over tapes	80	74	80	78
Signal 2,000 cps				
Tape 1	72	72	44	63
2	68	80	72	73
3	88	64	58	70
4	78	69	76	74
Mean over tapes	77	71	62	70

was presented in both temporal intervals. Since the signal was the same in both intervals, it cannot provide the cue for a choice between the intervals.

4. Complex signals, generated by adding four sinusoids together, were also tried as the signal in the percentage agreement analysis.

5. The background level of the noise was raised and lowered over a 40-decibel range to determine whether the absolute level of the background noise had an appreciable effect on the observer's consistency. This was done both with noise as a signal and for some sinusoidal-signal conditions.

None of these special presentation modes produced percentage agreement scores that appeared, on the basis of small samples, to be very different from those already reported in detail. In general, there was a small but consistent trend for tapes using sinusoidal signals to generate somewhat higher percentage agreement scores than tapes in which the signal was an increment in the power of noise. That the stimulus situation does not heavily influence the percent agreement scores suggests that the percentage agreement score may be largely determined by factors other than stimulus. Let us consider this alternative in more detail.

RESPONSE DEPENDENCIES

Because dependencies among the observer's responses can affect the percentage agreement score, we must discuss this dependency in some detail and try to estimate the magnitude of its effect in these experiments. The logical starting point is the simplest response dependency that one can encounter in a two-alternative task, namely, a pure response bias.

Pure Response Bias

If the observer always selected the *same* interval on all trials, his percentage agreement score, as analyzed from the no-signal trials would be 100%. To the extent that any observer increases the relative frequency of one response over another, the percentage agreement score is bound to increase, the amount of increase is easily determined. If the subject chooses the first interval of some no-signal trial with probability p, then he chooses the second interval with probability $q = (1 - p)$. Assuming independence between the two presentations the probability of agreement is simply the sum of these probabilities of agreement on the first interval p^2 plus the probability of agreement on the second interval q^2. That is, if only the response bias is operating and the subject chooses interval 1, 60% of the time, then his mean percentage agreement score is 52% $[(.6)^2 + (.4)^2]$.

As this example indicates, it takes a sizable bias to generate any appreciable effect on the percentage agreement score. In fact, if the bias, b, is taken to be the excess of the most popular interval over $\frac{1}{2}$, then the increase in the probability of agreement is $2b^2$. Thus for the percentage agreement score to exceed 60% on the basis of a response bias alone the selection of the most popular interval must exceed 72%. The largest pure response bias for any condition of the experiment was a selection of the first interval 62% of the no-signal trials. It is therefore not surprising that no significant percentage agreement score was found in any of the data collected.[11]

There is a variety of other simple response biases that one might consider, for example, alternation and double alternation, but such response chains can produce large percentage agreement scores only if the sequence starts on the same trial during the first and second presentation of the tape. In short, the start of the response chain must be synchronized to begin at the same place during both presentations. Thus pure response dependencies may have some influences on percentage agreement scores but their effects are certainly small. Let us consider a more likely candidate.

---

[11] It should be remembered that response bias was carefully monitored in the course of the experiments and the observers cautioned about over-frequency responses to one or the other intervals.

Stimulus-Response Dependencies

The occasion of the signal on the signal trials, or the subject's correct or incorrect response to that signal, does provide the synchrony needed to trigger response sequences that could generate agreements among the subject's responses on the successive presentations of the tapes. A careful analysis of the response following signal trials was, therefore, undertaken. Only one consistent pattern emerged. There was a tendency among all observers to choose the interval opposite the one on which they had just been correct. The effect was not large (60% at most) but was fairly consistent.

Such an effect, by itself, will not seriously affect percentage agreement, partly because, like a pure-response bias, the extent to which the bias affects the percentage agreement score is proportional to b^2 rather than b.

Furthermore, the bias observed, namely, to choose the interval opposite the interval that had just been correctly chosen, should show a variation with the percentage of correct detections on the signal trials. The results of Experiment I did not indicate such a trend.

There are many more analyses that might be undertaken in an effort to abstract response dependencies from the data.[12] One then might attempt to calculate their probable effect on the data obtained. A much more positive approach to this question would be to conduct the following experiment.

Crucial Experiment

The basic question is: How much of the percentage agreement score is the result of response dependencies, and how much is the result of stimulus factors? Any percentage agreement score reported above is an upper bound on the operation of stimulus factors because each score could be the result of response dependencies alone. The obvious way to investigate this problem is to generate a tape in which exactly the same stimulus occurs on both intervals of the no-signal trials. That is, the 50 no-signal trials might consist of 50 different samples of noise, but each noise sample would occur in pairs, the same wave form occurring on the first and second intervals of the forced-choice trial. Thus on a no-signal trial there would be not only no signal but no stimulus difference. Any percentage agreement score greater than chance would indicate the operation of response tendencies alone. It is the author's guess that the percentage agreement scores will be near chance under these conditions, but this is only conjecture, since admittedly any of the percentage agreement scores reported above could be the result of some intricate and undiscovered response chain.

Constructing such a tape would be most easily accomplished if a computer were available to generate the stimuli. One could also easily produce tapes in which some of the no-signal trials contain two different random samples of noise, whereas other trials contain identical noise samples. Then the relative consistency of the observer's behavior resulting from stimulus factors could be directly compared with those resulting from response factors. This experiment would also serve to clarify the amount of internal or biological noise present in these experiments.

---

[12] Published data on this problem does not give very clear advice on what bias one might expect. At last count, there were five studies reporting a tendency to alternate response alternatives, whereas three other studies reported that the subject tended to repeat the same alternative more than one would expect on the assumption of independence.

ESTIMATES OF INTERNAL NOISE

If we assume that all of our agreement scores are the result of stimulus factors, then we can estimate from the lack of agreement the magnitude of internal noise. Our estimates of internal noise are minimal estimates. Should the response dependency account for a large part of the agreement score, our estimate of the influence of stimulus factors would have to be reduced and our estimate of the amount of internal noise increased.

Estimates of internal noise can be obtained on the basis of the following model. The process of the observer listening to a no-signal trial can be likened to drawing two samples from a Gaussian population. Let the samples representing the two intervals be denoted x_1 and x_2. We assume that both have zero mean and variance, σ_E^2. This variance represents the variability of the external noise. The simplest assumption is that the internal noise sample, which we denote y_i, is added to the external noise.

The following simple decision rule is assumed,

If $x_1 + y_1 > x_2 + y_2$,
then choose interval one;
If $x_1 + y_1 < x_2 + y_2$,
then choose interval two.

The first sample of internal noise, y_1, should be independent of the second sample y_2, and we denote the variance of y_i as σ_I^2.

When the tape is presented the second time, x_1 and x_2 again occur as the samples from the taped external noise. These samples are now combined with new samples of internal noise; let us denote the new samples y_1' and y_2'. Once again the same decision rule is assumed, except y' replaces y.

TABLE 3

PROBABILITY OF AGREEMENT $P(A)$ VERSUS RATIO OF STANDARD DEVIATION OF EXTERNAL AND INTERNAL NOISE

$P(A)$	σ_E/σ_I	$P(A)$	σ_E/σ_I
0.503	0.10	0.803	2.10
0.512	0.20	0.811	2.20
0.526	0.30	0.818	2.30
0.544	0.40	0.824	2.40
0.564	0.50	0.830	2.50
0.585	0.60	0.836	2.60
0.607	0.70	0.842	2.70
0.628	0.80	0.847	2.80
0.648	0.90	0.851	2.90
0.667	1.00	0.856	3.00
0.684	1.10	0.875	3.50
0.701	1.20	0.890	4.00
0.716	1.30	0.901	4.50
0.730	1.40	0.911	5.00
0.743	1.50	0.925	6.00
0.755	1.60	0.936	7.00
0.766	1.70	0.944	8.00
0.777	1.80	0.950	9.00
0.786	1.90	0.955	10.00
0.795	2.00	0.982	25.00
		0.995	100.00

Agreement with the previous decision will be obtained if the inequality is in the same direction on both presentations.

Table 3 shows the probability of agreement for several ratios of external-to-internal noise.[13]

Since the average percentage agreement is nearly 66%, this model would suggest as a first approximation that the internal variability is nearly equal to the external variability.

Before we comment further on this estimate of the amount of internal noise, we should recognize the existence of a completely different set of assumptions which might also be used to account for the fact that the observer is not completely consistent. Rather than assume that some sensory or decision device is noisy or unstable, we might assume, instead, that the system has finite resolution. That is, we might assume that the decision device cannot distinguish between two sensory magnitudes unless they differ by some minimal amount. Such a quantal assumption would explain the consistency data reported in this paper. Whether this assumption is tenable in general would depend on the exact assumptions that one made concerning the nature of the processes. Certainly internal noise and quantal effects are often difficult to disentangle. The separation of these two hypotheses, which are very different in spirit at least, remains a challenging and important problem.

---

[13] The calculations shown in Table 3 were made at the Computation Center, Massachusetts Institute of Technology, on an IBM 7090 computer. The author wishes to acknowledge assistance from the Computation Center, Massachusetts Institute of Technology, and from Peter Reich who programed the computation.

We return to our estimates of internal noise based on the linear model. What are we to make of this estimate? There are certainly a good many assumptions in the model used to generate the calculations. The estimate is also only a lower bound. By itself this estimate of the amount of interval noise would be highly suspect. The surprising, and somewhat startling fact is that this paper represents the third method used to estimate this ratio and that all three methods yield estimates of the ratio very near one.

Swets, Shipley, McKey, and Green (1959), in determining the improvement from repeated observations of the same stimulus, estimate that the internal variability of the subject must be approximately the same as the external variability. Their estimates of σ_E/σ_I range between 1.0 and 1.4.

Watson measures a ratio similar to σ_E/σ_I by comparing what he calls an individual versus a group receiver operating characteristic curve. He estimates the value of this ratio is nearly one.

Whether this ratio is actually measuring the relative amount of internal and external variability is still uncertain. The consensus clearly suggests an interesting invariance in some aspect of the psychoacoustic situation.

CONCLUSIONS

The results show that the estimate of the ratio of internal-to-external variability is approximately one, and that this ratio is essentially constant over a wide variety of experimental conclusions. This estimate is in close agreement with similar estimates made by other investigators who studied both amplitude discrimination and the detection of a 1,000-cps sine wave in noise. This paper extends the range of experimental conditions by studying sinusoids of various frequencies, as well as signals that are samples of noise.

Another way of looking at this invariance is to note that the observer repeats his previous choices on approximately 70% of the two-choice trials. This result has important implications for the development of a molecular psychophysics, a psychophysics that attempts to predict the trial-by-trial behavior of the observer. This 70% consistency, since it appears to be independent of the sensory task, represents an upper bound on any attempt to predict trial-by-trial behavior. It is, of course, possible that some still undiscovered variable may permit us to raise this ceiling, but at present it remains the highest that can be achieved by a realistic simulator of an individual observer, namely that same observer on another occasion.

Certainly this comparatively low consistency score will dampen the enthusiasm of those who might wish to construct detection devices that will perfectly mirror the subject's trial-by-trial behavior. Indeed, not only is the consistency low, but the source of this rather meager degree of consistency is somewhat ambiguous, for it remains to be seen whether this 70% score is largely caused by stimulus or response processes. A simple extention of the present technique is proposed that should clarify this issue. However, even this rather low degree of consistency does not entirely preclude a molecular approach.

Throughout this paper, only an extreme form of molecular psychophysics has been considered, a psychophysics in which the level of analysis was the individual response. One might, however, present the same pair of wave forms 20 or 30 times, and measure the "most common" responses to the pair. The subject matter of such studies would be the relation between the acoustic wave form and this "most common" response.

The consistency of this "most common" response could be measured by a simple extension of this technique which was used to estimate the consistency of individual responses. Such extensions would provide interesting information about the manner in which the number of individual judgments used to determine the "most common" response affect the estimated variability of the internal factors.

Admittedly, the use of this most-common-response technique can never be as efficient as a trial-by-trial analysis, but comparing the prediction of two theories on this quasi-molecular basis may be considerably more efficient than trying to distinguish between molar predictions of threshold intensities that are only one decibel apart.

The development of some form of molecular psychophysics seems as inevitable as the development of more quantitative theories of sensory functions. Indeed, more and more crucial tests of such theories will be possible on the molecular level as they become more exact and quantitative. Detailed analyses of the determinants of the individual responses are necessary if we are to make reasonable and sensible use of such molecular data.

REFERENCES

Arons, L., & Irwin, F. W. Equal weights and psychophysical judgments. *J. exp. Psychol.*, 1932, **15**, 733–751.

Fernberger, S. W. Interdependence of judgments within the series for the method of constant stimuli. *J. exp. Psychol.*, 1920, **3**, 126–150.

Goodfellow, L. D. A psychological interpretation of the results of the zenith radio experiments in telepathy. *J. exp. Psychol.*, 1938, **23**, 601–632.

Luce, R. D. Individual choice behavior. New York: Wiley, 1959.

McGill, W. J. Serial effects in auditory threshold judgments. *J. exp. Psychol.*, 1957, **53**, 297–303.

Preston, M. G. Contrast effects and the psychometric judgments. *Amer. J. Psychol.*, 1936, **48**, 389–402. (a)

Preston, M. G. Contrast effects and the psychometric functions. *Amer. J. Psychol.*, 1936, **48**, 625–631. (b)

Senders, V., & Soward, A. Analysis of response sequences in the settings of a psychophysical experiment. *Amer. J. Psychol.*, 1952, **65**, 358–374.

Sherwin, C. W., Kodman, F., Jr., Kovaly, J. J., Prothe, W. C., & Melrose, J. Detection of signals in noise: A comparison between the human detector and an electronic detector. *J. Acoust. Soc. Amer.*, 1956, **28**, 617–622.

Skinner, B. F. The processes involved in the repeated guessing of alternatives. *J. exp. Psychol.*, 1942, **30**, 495–503.

Skinner, B. F. 'Superstition' in the pigeon. *J. exp. Psychol.*, 1948, **38**, 168–172.

SOLOMON, R. L. A note on the alternation of guesses. *J. exp. Psychol.*, 1949, 39, 322–326.

SPEETH, S. D., & MATHEWS, M. V. Sequential effects in the signal-detection situation. *J. Acoust. Soc. Amer.*, 1961, 33, 1046–1055.

SWETS, J. A., SHIPLEY, ELIZEBETH F., MCKEY, MOLLY J., & GREEN, D. M. Multiple observations of signals in noise. *J. Acoust. Soc. Amer.*, 1959, 31, 514–521.

WATSON, C. S. Signal detection and certain physical characteristics of the stimulus during the observation interval. Doctoral dissertation, University of Indiana, 1962.

(Received June 25, 1963)

Psychological Review
1964, Vol. 71, No. 5, 408-411

A THEORY OF CLASSICAL CONDITIONING

IN-MAO LIU
National Taiwan University

In view of the logical reasons and empirical evidences, the distinction between classical and instrumental conditioning is made in terms of experimental control. The stress on the nature of experimental control naturally leads us to identify 2 major components of conditioning. The function of the first component depends on the transfer of energy release (conditioning of autonomic responses) responsible for the initiation of a skeletal CR. Without operation of the second, the instrumental component defined by the effect of CR on UCS, the acquisition of CR is transient.

Classical and instrumental conditioning experiments refer, respectively, to training procedures in which the subject's response does not or does change the UCS. The reference experiment for classical conditioning is the study of conditioned salivation in dogs by Pavlov. Other experimental procedures which have also come to be called conditioning derived from the work of Bekhterev and Thorndike. The latter methods differ from the classical Pavlovian procedures in that the subject's behavior is instrumental to the production of reward or avoidance of punishment. Hence the name instrumental conditioning.

However, as was pointed out by Kimble (1961), it is difficult to produce pure cases of classical or instrumental conditioning. On the one hand, it is impossible to perform an instrumental-conditioning experiment without also providing the conditions for a classical one. Since a reward or punishment is an essential part of an instrumental-conditioning experiment, the situational cues regularly associated with reward or punishment function as conditioned stimuli. Similarly, unless the UCS precedes the CS in a classical conditioning experiment, the CR usually has some effect on the UCS. For instance, in eyelid conditioning, the occurrence of an anticipatory CR partially avoids the following airpuff and diminishes its intensity, introducing an instrumental component into the learning which takes place in a typical classical procedure. The manifestation of this instrumental component appears in the latency data of CR employing various interstimulus intervals. In Spooner and Kellogg's study (1947) of finger-withdrawal conditioning, the mean latencies of the CR for groups conditioned at CS-UCS intervals of .5, 1.0, and 1.5 seconds were, respectively, .419, .750, and 1.040 seconds. In this experiment as classical procedure was used and the UCS (shock) always followed the CS, it may be that the anticipatory occurrence of the CR acquired the secondary rewarding property through the subject's past life-long learning. The last statement suggests a possible way of understanding how the CR may actually be modified instrumentally in classical conditioning. In the history of the organism, the termination of painful stimuli has probably consistently been preceded by responses resembling the CR and, thus, in the manner of instrumental conditioning rewarded. This leaves in the subject a tendency to respond in ways which bring the reaction close to the occurrence (in previous life, close to the termination) of the UCS. Thus, even in GSR conditioning and heart-rate conditioning where one would be very hard put to show that the CR has its effect on the UCS, the response is part of a reaction which normally facilitates the mobilization of energy in an emergency situation. The result usually is to put an end to the noxious stimulation. The situation is no different from that in other learning situations involving a noxious UCS.

NATURE OF EXPERIMENTAL CONTROL IN CLASSICAL AND INSTRUMENTAL CONDITIONING

In the foregoing, an attempt was made to demonstrate the fact that in a typical Pavlovian conditioning experiment, an instrumental component of CR is always operating. Often it affects the value of the UCS, where it does not. Previous learning, where such an effect did occur, always seems possible to demonstrate. A consideration of the problem thus raised suggests a new way of looking at the nature of conditioning, the general nature of which we turn to now.

In the study already cited, Spooner and Kellogg (1947) also investigated backward and simultaneous conditioning. According to these results, the progress of conditioning shows an initial rise followed by a decrease to near zero toward the end of training. In all the instances there is one feature in common. The CR has no effect on the UCS. The effect here refers to both the primary and the secondary ones. This fact, together with the nature of the results, suggests that the mere contiguity of CS and UCS is sufficient to initiate the CR, but that some other process is necessary to make the CR persist once its establishment by contiguity has begun. This latter function is apparently effected by the instrumental aspect of the CR. Making this distinction we will refer to the process which depends upon the contiguity of CS and UCS as the initiation component and the other process as the instrumental component.

In what is usually called classical conditioning, the initiation component is under the direct control of the experimenter. However, the second, instrumental component, depends upon the subject's behavior, since the avoidance or weakening of the UCS depends on the amplitude and the other characteristics of the CR which are not under the direct control of the experimenter.

In what is usually called instrumental conditioning the instrumental component is directly under the control of the experimenter. The amount of reward or punishment is controlled by the experimenter and delivered for a range of responses which differ markedly in their details. The initiation component, however, is not directly specifiable in a usual instrumental-conditioning experiment, because the time interval between the presentation of the discriminandum (CS) and the occurrence of reward or punishment (UCS) is a variable depending on the details of the subject's behavior.

From the preceding analysis, it follows that the usual treatments of classical and instrumental conditioning emphasize only one of the two essential components of the conditioning process.

INITIATION AND INSTRUMENTAL COMPONENTS OF CONDITIONING

The two-component theory of conditioning developed above holds that the initiation component is responsible for the first occurrence of CR and that the instrumental component accounts for the later firm establishment or change in the CR. The implications of this analysis are most significant for classical conditioning. To develop some of these implications, we start by analyzing the concept of UCS.

In the first place the UCS is a stimulus and therefore capable of arousing a certain sensory process within the organism, and the more important property of the UCS is that it regularly evokes a response, the UCR. The regularity with which the UCS evokes a response of certain strength, in one way of thinking, reflects a drive function of the UCS. The drive function as used here refers only to that function which relates stimulus to response. Paired presentations of the CS and the UCS (conditioning) thus may be viewed as a means of producing a transfer of drive function to the CS.

The transfer of drive function presumably depends on the transfer of energy mobilization. In other words, the conditioning of autonomic responses underlies and is indispensable for the conditioning of a skeletal response. By means of the conditioning of autonomic responses energy is made available for the initiation of the CR to an originally neutral stimulus, CS. It is well known to a student of learning that success in conditioning a reflex has been notably better in experiments employing electric shock, though the reflex may be readily elicited by some other stimulus which does not evoke autonomic responses.

The concept of reward such as measured by the weight of food is probably not a significant variable which psychologists are looking for. Neither is it in accordance with the concept of drive reduction nor with the idea of fear conditioning. They lack power to encompass a variety of behavioral phenomena. The transfer of energy availability is assumed to be basic to the initiation of a CR to a neutral stimulus. A brief excitation with the subject provides the basis for the transfer of energy availability in a positive reward training as well as in a situation involving a noxious stimulation. The initiation component thus directly depends on the transfer of energy availability.

Therefore, a close relationship between the CS-UCS interval function of a skeletal response and that of heart beat should be expected. As a matter of fact this has already been pointed out by Church and Black (1958). Namely, the latency of the first avoidance response, as a function of the CS-UCS interval, coincides fairly well with the mean latency of maximum heart beat.

When the magnitude of the UCR is directly related to the amount of energy mobilization, the correlation between the magnitude of the UCR and that of the CR should, according to the present analysis, be nearly perfect. Such is the case as obtained in the study of heart-rate conditioning by Black, Carlson, and Solomon (1962).

How the relationship between the conditioning of autonomic responses (especially those which are connected with energy release) and that of skeletal responses is established through the ontogenetic development is a challenging problem. In the course of the ontogenetic development, whenever there is a modification of behavior, there is conditioning of those autonomic responses responsible for energy release. This means that the conditioning of those autonomic responses is much more frequent than that of specific skeletal responses. Accordingly, the conditioning of those autonomic responses should be very rapid, because we know that repetition of conditioning accelerates the rate of conditioning. Empirically, the acquisition and extinction in heart-rate conditioning are within 5 trials. This contrasts with the rate of conditioning of eyeblink or finger withdrawal, where more than 80 trials are usually required for acquisition to an asymptotic level.

Another implication is that those autonomic responses which mediate energy release can be conditioned with backward procedure. Since the experimental evidence indicates that the conditioned skeletal response established by backward conditioning procedure is transient, those autonomic responses, in order to supply necessary energy for the initiation of the skeletal CR, should be conditioned beforehand. There is no empirical study on heart-rate conditioning using backward procedure. However, an evidence is available for the backward conditioning of GSR (Jones, 1962).

We turn now to the nature of the instrumental component. As stated in the previous section, without the operation of the instrumental component, the initiated response is transient. In this sentence, some inhibitory function of the UCR on the CR, when the CR does not affect the UCR, is assumed. At this point, a question may be raised to the fact that heart-rate conditioning can be obtained in totally curarized dogs (Black, Carlson, & Solomon, 1962); even the CR (increase in heart rate) cannot very easily alter the effectiveness of the UCS. The answer is in terms of frequency theory. The secondary effect of heart-rate conditioning in terminating noxious stimuli by means of specific skeletal responses has occurred so frequently in the past history of the organism that the CR (increase in heart rate) has become more or less persistent; even it does not affect the UCS. This also explains the wide range of the CS-UCS interval function in heart-rate conditioning. Consider the reverse to be true, then the conditioning of a specific skeletal response is scarcely possible, because not all skeletal responses terminate noxious stimulation automatically from the beginning, especially in the usual instrumental conditioning.

REFERENCES

BLACK, A. H., CARLSON, N. J., & SOLOMON, R. L. Exploratory studies of the conditioning of autonomic responses in curarized dogs. *Psychol. Monogr.*, 1962, **76**(29, Whole No. 548).

CHURCH, R. M., & BLACK, A. H. Latency of the conditioned heart rate as a function of the CS-US interval. *J. comp. physiol. Psychol.*, 1958, **51**, 478–482.

JONES, J. E. Contiguity and reinforcement in relation to CS-UCS intervals in classical aversive conditioning. *Psychol. Rev.*, 1962, **69**, 176–186.

KIMBLE, G. A. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

SPOONER, A., & KELLOGG, W. N. The backward-conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321–334.

(Received August 22, 1963)

ERRATUM

In the article by Warner R. Wilson and Howard Miller in the May 1964 issue, line 3 of column 2 on page 238 should read,

compatible with r-s since in r-s one

and on page 241, column 2, line 9 of the first paragraph should read,

a-s-design (a-s-d) and to the some-dif-

Psychological Review
1964, Vol. 71, No. 5, 412–414

THEORETICAL NOTE

VERBAL TRANSFER AND INTERLIST SIMILARITIES[1]

JOHN P. HOUSTON

University of California, Berkeley

Degree and direction of verbal transfer were related to first list stimulus–second list response similarity (S_1-R_2) and to second list stimulus–first list response similarity (S_2-R_1). This was accomplished through a relabeling of Osgood's transfer and retroaction surface such that the dimensions referred, not to degree of S_1-S_2 and R_1-R_2 similarity, but to degree of S_1-R_2 and S_2-R_1 similarity. When S_2-R_1 similarity is minimal transfer will be minimal, no matter what the degree of S_1-R_2 similarity. As S_2-R_1 similarity increases transfer increases until, when S_2 and R_1 are identical, maximum negative transfer will obtain when S_1-R_2 similarity is minimal and maximum positive transfer will occur when S_1 and R_2 are identical.

Although the importance of the similarity variable in relation to transfer phenomena has long been recognized there has been a tendency among psychologists to focus upon particular loci of similarity and to ignore certain others. The sets of similarity relationships which have received the most attention in the typical two list S_1-R_1, S_2-R_2 verbal transfer situation are those expressed by Osgood's (1949) transfer and retroaction surface. His three dimensional surface represents an attempt to relate direction and degree of transfer to first list stimulus–second list stimulus similarity (S_1-S_2) and to first list response–second list response similarity (R_1-R_2).

Two loci of similarity which have been virtually ignored in the study of transfer may be described as the similarity between first list stimuli and second list responses (S_1-R_2) and the similarity between second list stimuli and first list responses (S_2-R_1). What will be the influence, if any, of varying degrees of S_1-R_2 and S_2-R_1 similarity upon direction and degree of transfer? Although the verbal learning literature contains a number of examples discussed below which suggest what these influences might be, there has been no attempt to provide a general integrative statement of the relationships between transfer and the S_1-R_2 and S_2-R_1 similarity loci. The intent of the present paper is to provide this kind of a statement, based upon data, through a relabeling of the dimensions of Osgood's transfer surface such that the surface corresponds, not to the amount of transfer associated with degrees of S_1-S_2 and R_1-R_2 similarity, but to the amount and direction of transfer associated with varying degrees of S_1-R_2 and S_2-R_1 similarity. Figure 1 contains the relabeled surface.

As is true of Osgood's surface the vertical dimension represents direction and degree of transfer. The length of the horizontal plane, labeled S_1-R_2, represents degree of similarity between first list stimuli and second list responses. The width of the horizontal plane, labeled S_2-R_1, represents degree of similarity between second list stimuli and first list responses. In both cases the similarity dimension extends from maximum similarity or identity (labeled 1) to minimum similarity or neutrality (labeled 0). If the first list responses are different from and neutral with respect to the second list stimuli then, no matter what the de-

---

[1] This paper was written during the tenure of the author's United States Public Health Service postdoctoral fellowship at the Institute of Human Learning, University of California, Berkeley.

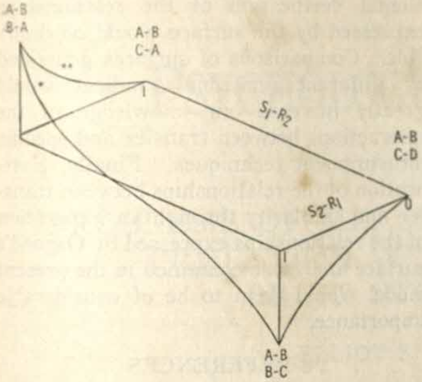


FIG. 1. Transfer as a function of S_1-R_2 and S_2-R_1 similarity.

gree of second list response–first list stimulus similarity, transfer will be minimal. As the degree of S_2-R_1 similarity increases transfer will increase until, when S_2 is identical to R_1, maximum positive transfer will obtain when R_2 is identical to S_1 and maximum negative transfer will occur when R_2 and S_1 are different and neutral.

The surface is not completely unsubstantiated. Although little attention has been given to S_1-R_2 and S_2-R_1 transfer effects per se, the verbal learning literature does contain some data which document a number of points on the surface. Murdock (1956) found that A-B learning will facilitate subsequent B-A learning. His experimental paradigm corresponded to the point of maximum positive transfer on the proposed surface where S_1 is identical to R_2 and S_2 is identical to R_1. Murdock explained this instance of positive transfer in terms of B-A strength established during the A-B learning stage. In another study Murdock (1958) has confirmed an additional point on the surface. He employed an A-B, B-C paradigm and found a negative transfer effect during second list learning. This situation corresponds to the point of maximum negative transfer where S_2 and R_1 are identical and R_2 and S_1 are different and neutral with respect to each other. Murdock proposed that the B-A connection established during A-B learning will compete with the B-C learning during the second stage.

An experiment by Young (1961), while not conforming to the usual two list transfer situation, produced some pertinent data. He constructed lists such that all items served as both stimuli and responses and found that these lists were more difficult to learn than lists whose items did not serve in the double capacity. The proposed transfer surface would predict this result. His lists were, in effect, composed of B-C pairs preceding A-B pairs and A-B pairs preceding B-C pairs. According to the surface the former sequence of pairs should produce no transfer while the latter should produce negative transfer effects through the competition of the B-A associations with the B-C associations. Thus slow overall learning would have been predicted for the double-function list.

Further documentation of the surface may be found in the field of verbal mediation. Although not strictly applicable to the similarity surface contained in Figure 1, as some of the relationships are associative, these data are at least suggestive of the kinds of results which might be obtained with similarity measures. Kjeldergaard and Horton (1962) employed two mediation paradigms which may be located on the transfer surface. The first was a C-B, B-A, A-C paradigm. If one thinks of this situation as a two stage B-A, A-C transfer paradigm with the C-B list establishing a relationship (in this case associative) between the first-stage stimuli (B) and the second-stage responses (C) then it corresponds to a transfer situation with maximum S_2-R_1 similarity (A and A) and some positive amount of S_1-R_2 relatedness (B-C). In accordance with the transfer surface prediction for such a paradigm (see * in Figure 1) Kjeldergaard and Horton found A-C facilitation or positive transfer. The second relevant paradigm was an A-B, C-B, A-C situation. Once again, if A-B is interpreted as a definition of a positive relationship between the first-stage response (B) and the second-stage stimulus (A) then the paradigm corresponds to a transfer situation in which

S_1-R_2 (C-C) is maximum and S_2-R_1 (A-B) is some positive value (see ** in Figure 1). Kjeldergaard and Horton reported A-C facilitation. The transfer surface predicts positive transfer if A-B strength is assumed to be of some minimal degree.

Thus there is some evidence or support for a number of points on the surface. Of course, the majority of the relationships depicted by the surface have never been tested and require empirical verification, particularly under constant experimental conditions.

The surface is intended as nothing more than a rough index of the relationships existing between transfer and S_1-R_2 and S_2-R_1 similarity. Presumably varying materials, experimenters, and situations will produce variations in the shape of the surface. Qualitative as well as quantitative changes in the relationships between stimuli and responses should also influence the shape of the surface. Thus, formal similarity, judged similarity, associative strength, synonymity, and associative overlap may each generate surfaces which, while conforming to the general characteristics of the Figure 1 surface, may differ from each other considerably.

The proposed surface, as a statement of the relationships between transfer and two specific similarity loci, suggests a number of directions for further study. As mentioned, a good deal of experimental verification of the relationships expressed by the surface would be desirable. Comparisons of surfaces generated by different relatedness indices would greatly increase our knowledge of the interactions between transfer and specific measurement techniques. Finally, clarification of the relationships between transfer and similarity through an integration of the relationships expressed by Osgood's surface and those contained in the present model would seem to be of considerable importance.

REFERENCES

KJELDERGAARD, P. M., & HORTON, D. L. An experimental analysis of associative factors in stimulus equivalence, response equivalence and chaining paradigms. *Studies in verbal behavior.* Report No. 3, National Science Foundation Grant, University of Minnesota.

MURDOCK, B. B. "Backward" learning in paired associates. *J. exp. Psychol.,* 1956, **51,** 213–215.

MURDOCK, B. B. "Backward" associations in transfer and learning. *J. exp. Psychol.,* 1958, **55,** 111–114.

OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.,* 1949, **56,** 132–143.

YOUNG, R. K. Paired-associate learning when the same items occur as stimuli and responses. *J. exp. Psychol.,* 1961, **61,** 315–318.

(Received December 9, 1963)

VOL. 71, NO. 6 NOVEMBER 1964

PSYCHOLOGICAL REVIEW

PROBLEMS OF MEASUREMENT AND INTERPRETATION WITH REINFORCING BRAIN STIMULATION[1]

ELLIOT S. VALENSTEIN

Fels Research Institute

An increasing number of experiments have been concerned with various aspects of reinforcing brain stimulation. The dramatic nature of many of the findings, however, has often forced methodological and interpretive problems into the background. The present paper discusses some of the problems encountered in: (a) measuring the reinforcement strength, (b) determining stimulus thresholds, (c) interpreting interactions between specific neural areas.

The discoveries that electrical stimulation of the brain may have positive and negative reinforcing properties have triggered experimentation and speculation which have been increasing at a rapid rate.[2] The possibility that the road was now open to the discovery of basic physiological mechanisms of motivation and emotion aroused the active interest of psychologists, physiologists, pharmacologists, anatomists, and others. Because of the interdisciplinary nature of much of the research and in part because the dramatic nature of the findings tended to force questions of methodology into the background, many of the complexities of the methods adopted have not been fully appreciated. As there have been a number of attempts in recent years to integrate experimental findings into theoretical schemes, it would seem appropriate at this time to examine some of the problems of measurement and interpretation of the results obtained with reinforcing brain stimulation.

In emphasizing the problems of measurement and the determination of what has been measured there is no intent to restrict exploration, but rather to locate buoys which point out where there is a risk of running aground. If more attention is not given to these problems there is the danger of being overwhelmed by an accumulation of anecdotal reports which will seriously impede progress in this field.

Although there may be many ways of organizing a discussion of the problems of measuring and interpreting the reinforcing consequences of brain stimulation, the major issues can be conveniently, if somewhat arbitrarily, grouped under three headings: (*a*)

---

[1] The author wishes to acknowledge the support of Research Grants MH-4529 National Institutes of Health, NsG-437 National Aeronautics and Space Administration and Career Development Award MH-K6-4947 National Institutes of Health.

[2] For a comprehensive review the reader is referred to Olds, 1962.

measurement of reinforcement strength, (*b*) determining stimulus thresholds, and (*c*) interpreting interactions between specific neural areas.

MEASURING REINFORCEMENT STRENGTH

Following the initial demonstration that animals would respond in order to obtain electrical stimulation of certain brain areas, it soon became apparent that methods to quantify the strength of the reinforcement were needed. Studies of the relative reinforcement strength of stimulation of different neural areas and estimates of change in reinforcement value with different stimulus parameters, drugs, ablations, gonadectomy, and deprivation have been the subjects of numerous studies (see Olds, 1962). These studies all require methods of determining change in the reinforcement properties of brain stimulation.

The most widely used measure of the strength of a reinforcer is the rate of making some response (usually lever pressing) which is followed by the presentation of that reinforcer. In addition to the convenience of this measure there is much appeal in the argument that the length of time permitted to elapse between responses reflects the intensity of the desire for the reward. There are difficulties, however, of both an empirical and logical nature. Empirically, it has been shown that response rate can be a misleading index of reinforcement with brain stimulation as results may not agree with an animal's preference (Hodos & Valenstein, 1962) or with measures of resistance to competition from other reinforcers such as food and shock avoidance (Valenstein & Beer, 1962). At high intensities, for example, response rate usually declines because motoric side effects of the stimulation disrupt performance, but animals may choose those intensities over a lower amplitude stimulation that supports a higher response rate.

Hawkins and Pliskoff (1964) have confirmed these findings recently with a new technique. These authors have employed a two-member behavioral chain in which responses on the first lever were reinforced on a variable-interval schedule with the insertion of a retractable lever. Responses on the retractable lever provided brain-stimulation trains on a continuous reinforcement schedule. A comparison of the rates on the two levers revealed that response rate on the first lever continued to increase at stimulus intensities higher than those which produced peak response rates with continuous reinforcement. It was concluded that brain stimulation could not be assessed adequately by self-stimulation rates with continuous reinforcement.

Logically, there are also difficulties. When average response rate is used as a measure of the value of the reinforcer to the animal there exists an implicit assumption that reinforcement strength is homogeneous throughout the testing session. This assumption can not always be met with brain stimulation as some effect of the stimulus which persists may change the value of the reinforcer after its administration. The finding that seizure activity is frequently associated with reinforcing brain stimulation (Newman & Feldman, 1960; Porter, Conrad, & Brady, 1959) suggests that some aspect of the stimulation may persist and obviate the necessity to respond immediately. Equally likely is the possibility that a second stimulation may result in a qualitatively different experience if sufficient time has not elapsed.

Some data collected in our laboratory is pertinent to the point. It has been observed repeatedly that animals with septal electrodes characteristically

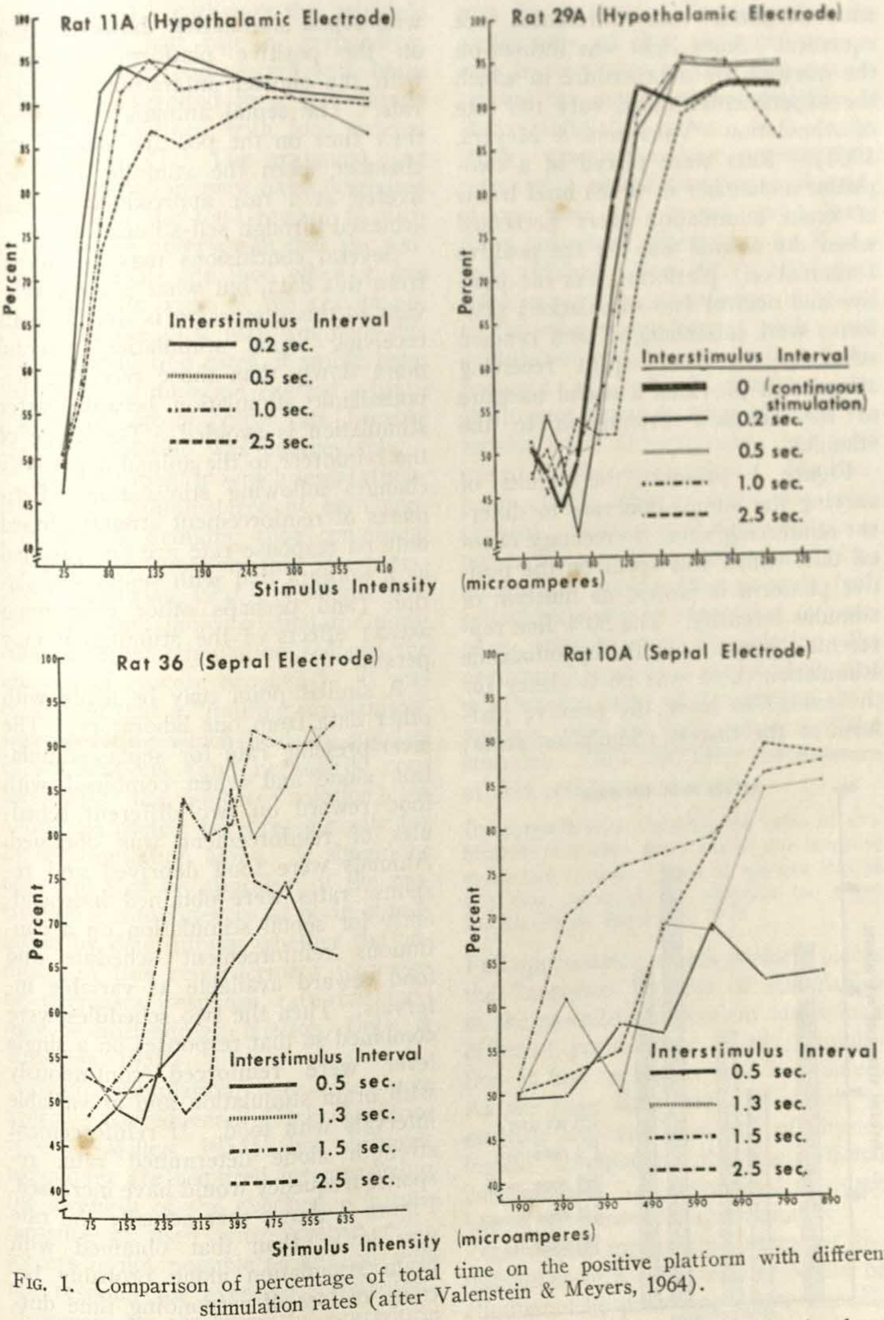


FIG. 1. Comparison of percentage of total time on the positive platform with different stimulation rates (after Valenstein & Meyers, 1964).

stimulate themselves more slowly than do those with electrodes in the posterior hypothalamus. With the self-stimulation procedure animals have control over the rate of stimulation and interpretations of the meaning of such response rate differences are equivocal. Some light was thrown on the question by a procedure in which the experimenter could vary the rate of stimulation (Valenstein & Meyers, 1964). Rats were placed in a two-platform chamber in which brief trains of brain stimulation were presented when the animal was on the positive (stimulation) platform. As the positive and neutral (no-stimulation) platforms were interchanged on a random schedule, the time spent receiving stimulation provided a useful measure of the animal's orientation to the stimulus.

Figure 1 presents the results of varying the stimulation rate to different reinforcing sites. Percentage of total time which was spent on the positive platform is plotted as function of stimulus intensity. The 50% line represents chance. With hypothalamic stimulation there was no tendency for the animal to leave the positive platform at the fastest stimulation rates; with septal stimulation, however, time on the positive platform decreased with the shorter interstimulus intervals. The septal animals maximized their time on the positive side of the chamber when the stimulus was delivered at a rate approximating that achieved through self-stimulation.

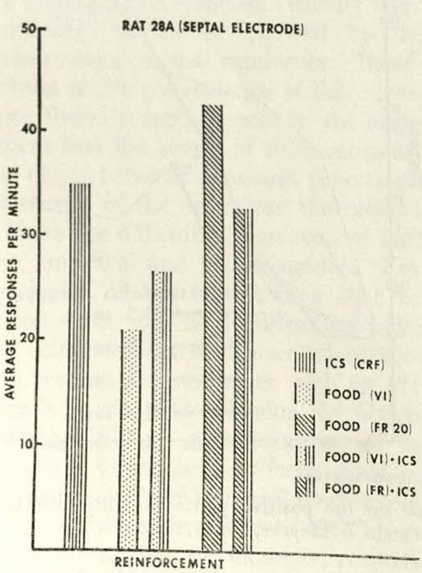


FIG. 2. Average response rate for brain stimulation and food reinforcement presented separately and together.

Several conclusions may be drawn from this data, but what is most relevant to this discussion is that animals receiving septal stimulation respond more slowly than those receiving hypothalamic stimulation because faster stimulation is avoided. The value of the reinforcer to the animal apparently changes following stimulation. Estimates of reinforcement strength based only on response rate are not justified as it appears that with septal stimulation (and perhaps other reinforcing areas) effects of the stimulation may persist.

A similar point may be made with other data from our laboratory. The lever-pressing rate for septal stimulation alone and when combined with food reward on two different schedules of reinforcement was obtained. Animals were food deprived and response rates were obtained independently for septal stimulation on a continuous reinforcement schedule and food reward available at variable intervals. Then the two schedules were combined so that responses on a single lever were reinforced continuously with brain stimulation and at variable intervals with food. If reinforcement strength alone determined rate, response frequency would have increased. Figure 2 shows that actually the rate was lower than that obtained with brain stimulation alone probably because of loss of responding time during eating. A more striking demonstration was obtained when the animal was placed on a schedule requiring 20 responses for each reinforcement (fixed ratio). Although this schedule generates a fast rate of responding, combining the food reward with brain stimulation resulted in a rate slower than that obtained with food reward alone (Fig. 2). The argument that brain stimulation may have decreased the value of the food reward is mitigated by the observation that the animal always ate the food when it was available. In view of the results obtained on the dual platform with different stimulation rates, it would seem that septal stimulation exerted a braking action on response rate because faster stimulation was unacceptable.

An experiment in which septal stimulation was administered at each contact with a drinking tube containing glucose also supports this interpretation (Asdourian, 1962). Changes from base-line data indicated that the addition of septal stimulation reduced the amount of glucose consumed. It was concluded that a "limiting mechanism" was operating which held the number of stimulations (and necessarily contact with the glucose solution) to about the pretest bar-pressing rate for brain stimulation alone. It should be apparent that an appraisal of the motivational consequences of brain stimulation by determining whether the addition of this reinforcement increases or decreases response rate for food may be misleading (Nielson, Doty, & Rutledge, 1958).

Brain stimulation may also have aversive consequences (Delgado, Roberts, & Miller, 1954), but different techniques are required for measuring the strength of the aversion. A rate measure as used with positive reinforcers is not possible as animals will not respond for aversive stimulation. However, by changing the procedure so that stimulation is postponed for a fixed period of time following each response, an avoidance (or escape) rate may be obtained. To determine the significance of this response rate, it may be necessary to take into consideration the matter of acceptability of faster stimulation rates. In a recent study, stimulation was presented in trains of .5-second duration at the rate of one train per second and a lever press postponed stimulation for 4 seconds (Olds & Olds, 1963). Response rates to postpone stimulation indicated an avoidance tendency, while self-stimulation performance indicated the presence of positive reinforcement. With pure negative reinforcement the results were unambiguous, but with mixed approach-avoidance (ambivalent) effects interpretation was complicated. Some septal electrode placements, for example, were classified as ambivalent because the animals both self-stimulated and responded to postpone stimulation. Considering the previous discussion it seems likely that with slower rates of stimulation these animals may not have terminated the stimulus. Olds and Olds were aware of this problem and wrote:

> In approach tests, the response rates of ambivalent rats were never above one response every two seconds. Thus it appears that in this case, applying the stimulus too often has avoidance effects [p. 277].

Full appreciation of this problem places the "aversive" aspects of stimulation in the so-called ambivalent areas in a different perspective. Only stimulation at fast repetition rate is avoided. As we have little basis by which to evaluate the repetition rate of physiological stimulation, the motivational consequences of activation of such neural structures is questionable.

It would appear that many of the difficulties of a rate measure would be eliminated by providing brain stimulation less frequently than with each response. With such methods, both the disruption of performance by motoric side effects and the preference for slower stimulation rates would have less influence on response rate. Indeed, the demonstration that intermittent brain stimulation was capable of maintaining behavior was accompanied by the suggestion that this would minimize the influence of gross motor effects of the stimulus on the response rate (Sidman, Brady, Boren, & Conrad, 1955). In view of this advantage of intermittent reinforcement schedules, it might seem strange that regular reinforcement continues to predominate. A partial explanation may be found in the difficulty of maintaining the behavior of many animals with only intermittent brain stimulation. Perhaps comparable is the observation that some animals which press a lever repeatedly for continuous reinforcement do not perform adequately in a simple maze or runway where reinforcement rate is less frequent (Newman, 1961; Olds, 1956; Spear, 1962). There may be a species difference in this respect, as monkeys have been shown to respond stably on fixed ratio schedules requiring 100 responses for one reinforcement (Brodie, Moreno, Malis, & Boren, 1960). In the past, however, brain-stimulation studies with rodents have used only schedules providing a high reinforcement rate which is in striking contrast to the routine testing with ratios of 100 or more with food reinforcement (Ferster & Skinner, 1957).

Recently, however, Pliskoff and Hawkins (1964), using the technique described above, have been able to obtain stable data from rats with schedules providing a low frequency of reinforcement. It will be recalled that these authors use a procedure in which animals are required to press one lever which is reinforced on some schedule with the insertion of a retractable lever. The animal may then receive a brain stimulation train for each response on the retractable lever up to some predetermined number of presses before this lever is withdrawn. With this procedure it has proven possible to maintain behavior on the first lever with variable-interval reinforcement schedules up to VI 4 minutes and fixed ratio schedules up to FR 200. Although this method does not appear equally successful with all reinforcing neural sites, it does suggest that the response rates on the first lever may provide very useful information for assessing the value of the stimulation.

Some caution in using response rate even with intermittent reinforcement should be observed. While response rate appears to reflect the incentive value of the stimulus, it may not similarly reflect internal states of the organism which would be expected to interact with reinforcement value. With food reward, for example, it has been shown that response rate maintained by a variable-interval schedule of reinforcement may be insensitive to deprivation level. Also following repeated testing with fixed ratio schedules, reinforcement may become associated not only with the response, but also with a particular rate of responding. Under these conditions, rate may become an essential part of the response and changes in reinforcement value may not be reflected in response rate (Sidman, 1960). Response rate develops a tempo or rhythm and becomes increasingly stable, but at the same time it is likely to become less sensitive as a measure. The paradox is that until there is a satisfactory independent measure of reinforcement strength, it will be impossible to determine how faithfully response rate reflects reinforcement value. With food reward, the assumption that reinforcement increases with deprivation or quantity of food has common sense appeal. With brain stimulation there are few convenient guides as manifested by conflicting opinion on the effect of increasing stimulus intensity on reinforcement value (Hodos & Valenstein, 1962; Reynolds, 1958).

Tests which permit an animal to demonstrate a preference may provide a relative measure of reinforcement strength. These tests may consist only of presenting two reinforcing conditions simultaneously and recording some index of the animal's preference. Quantification of the strength of the preference may be accomplished with techniques which require that the animal perform some work in order to change from the less preferred to the more preferred reinforcement. By adjusting the amount of work to a point of equal preference, an estimate of the degree of preference may be achieved. Verhave (1963) has described some of the experimental variables and mathematical considerations applicable to this type of testing procedure when used with food reward. In general, however, preference tests tend to become cumbersome when a large number of stimuli are to be compared. For most experimental purposes, it would be more convenient if reinforcement value could be expressed in terms of a metric unit that would indicate relative position on a scale. Preference tests, however, may help to validate a method which does provide such information.

Tests other than response rate may provide a useful unit of measure. One avenue which has been explored is based on a determination of how much an animal will overcome to obtain reward. Obstruction-box techniques (Warden, 1931) have been generally rejected because of the variability of behavior resulting from repeated electric shock. Recently a test was described in which the number of lever presses required for successive reinforcement increased by a fixed ratio (Hodos, 1961). The measure of reward strength resulting from such a "progressive ratio test" is the number of unreinforced responses an animal will make before the behavior is extinguished. This technique appears to be useful with food reward, but only preliminary information exists with reinforcing brain stimulation (Hodos, 1963). The method has the advantage of eliminating the problems inherent with a response-rate measure and also provides a unit useful for comparative purposes. It will be important, however, to determine if all self-stimulating animals perform reliably under conditions which provide only intermittent reinforcement.

Another measurement problem about which much interest has centered concerns the changes in reinforcement value as a function of the duration of stimulation. When given control over stimulus duration, animals will terminate positive stimulation. This has been attributed to the excitation of a neighboring aversive neural system through a temporal summation of inadequate stimuli (Stein, 1962a). The most frequently used method of obtaining preferred duration permits the animal to hold a single lever down to obtain stimulation which is terminated when the lever is released (Bower & Miller, 1958; Stein, 1962a). This method yields very brief "preferred" durations, but it would appear that the influence of motoric side effects of the stimulation have not been considered sufficiently. Very different data are obtained when animals are not required to hold the lever, but are free to move about the testing chamber until they press a second lever which terminates the stimulus. Direct observation of the behavior suggests that the different results are due to the in-

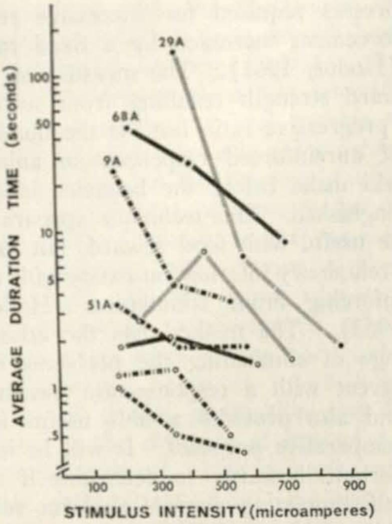


FIG. 3. Comparison of average duration animals maintain brain stimulation when obtained with one-lever (open circles) and two-lever (filled circles) techniques.

ability of the animals to inhibit motor reactions forcing them off the lever, which in the one-lever test terminates the stimulus.

Figure 3 contrasts the results from these two methods with four animals receiving reinforcing hypothalamic stimulation (Valenstein & Valenstein, 1963). It is evident that the preferred durations obtained by the two methods are of different orders of magnitude and are not to be explained away by the fraction of a second necessary to traverse the short distance between the levers. The interpretation suggested by one set of results does not seem to be appropriate to the other. The notion of activation of an aversive system through temporal summation may be appropriate to durations in the order of 1 second, but would require a considerable extension of the neurophysiological data to be applied to durations above 10 seconds.[3]

Speculation (Miller, 1957) that there exists a unitary neural system underlying all reinforcement which is constantly being modulated by separate drive states has been the impetus for much research. Those neural areas which animals will stimulate themselves are viewed as the best candidates for this unitary reinforcing system. The possibility that activation of drive states may alter the level of excitability of the neural tissue crucial to self-stimulation behavior is implicit in this theory. There is indeed some support for this idea as it has been reported that ingestion of nutrients may change the activity of cells in hypothalamic nuclei (Anand, Chhina, & Singh, 1962; Anand, Dua, & Singh, 1961). Food deprivation also has been shown to increase self-stimulation rate (Brady, Boren, Conrad, & Sidman, 1957; Hodos & Valenstein, 1960; Olds, 1958), but these studies raise a methodological question that is basic to experiments of this design. Figure 4 presents self-stimulation rates of animals when sated and hungry. It can be seen that there are striking differences under the two conditions, but the response rates are considerably below the level of 30 to 100 per minute commonly observed with reinforcing brain stimulation. It becomes necessary to ask whether the stimulation was actually reinforcing at the intensities used to demonstrate these differences. This question is especially critical in view of the well-established relationship between food deprivation and activity level. Activity changes occur not only with food deprivation, but with the administration of depressant and excitant drugs as well as with stimulation of some neural structures. For example, recently stimulation of the caudate nucleus of the cat (a suggested reinforcing site) has been shown to produce hyperactivity which with some testing procedures may be confused with reinforcement (Justesen, Sharp, & Porter, 1963). Before a rate change can be attributed to any direct effect on the neural substrate of reward, the influence of general activity level on performance must be parceled out as it has been shown that even in the absence of stimulation, food-deprived animals respond at significantly higher rates than sated animals (Hodos & Valenstein, 1960).

---

[3] Temporal summation is generally explained by the addition of excitatory postsynaptic potentials (EPSP). These potentials reach a peak in a few milliseconds and thereafter rapidly decay. Afferent subthreshold volleys, for example, are unable to generate an impulse unless the interval between volleys is less than 5 milliseconds (Eccles, 1957).

THRESHOLD DETERMINATION WITH REINFORCING BRAIN STIMULATION

There are several reasons why the determination of a stimulus intensity threshold is important with brain-stimulation experiments. Self-stimulation behavior would be expected to be most sensitive to influence at low intensities. Presumably, this is true because changes in the excitability of the nervous tissue are likely to have significant influence only if the stimulation intensity is not so high that it overcomes all differences in nerve cell responsiveness. The general acceptance of this point of view can be seen in Stein's (1962c) statements, "drugs effective against depression should increase the ability of the brain to respond to positive reinforcement, either by directly stimulating or sensitizing reward centers, or by selectively inhibiting aversion centers [p. 298];" and "It would be hard to exaggerate the importance of using threshold-intensity currents for these tests [p. 299]."

Threshold determination is also important for charting the reinforcing conseqences of stimulation of different neural areas as it is reasonable to assume that the threshold reflects prox-

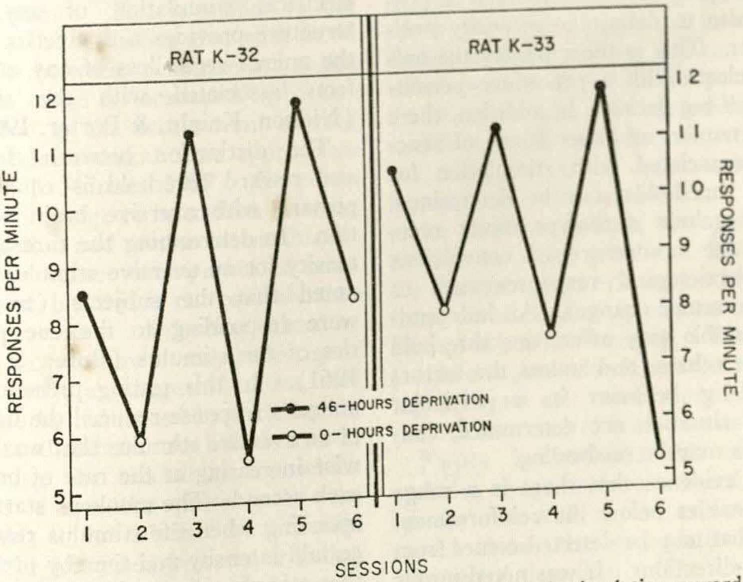


FIG. 4. Lever-pressing rates for intracranial electrical stimulation reward in the two rats as a function of deprivation interval (after Brady, Boren, Conrad, & Sidman, 1957).

imity to the reinforcing site. In addition, the threshold intensity provides some indication of the relative size of the neural field which must be excited in order to achieve effects.

Actually the stimulus threshold has several possible meanings and it is essential that these different meanings be kept separate. Proceeding from the low end of the intensity scale upward we may speak of the cue or detection property of the stimulus, and the lowest intensity which can be sensed, directly or via some mediating process, may be called the *detection threshold*. The animal perceives the stimulus and can be trained to respond to the brain stimulation as to any other discriminable stimulus, but prior to learning, the stimulus is neutral and possesses no special reinforcing consequences. At higher intensities we may speak of a *reinforcement threshold*, which is the minimum intensity capable of maintaining some instrumental behavior to obtain or escape from the stimulus. With appropriate techniques it is possible also to determine *intensity preferences*. This is the intensity the animal selects with a procedure permitting self-regulation. In addition, there are a number of other kinds of reactions associated with stimulation for which thresholds may be determined. These include stereotype motor reactions, the occurrence of convulsions and physiological responses such as cardiovascular changes. An independent variable may affect one threshold but not others, and unless the factors controlling behavior in a particular testing situation are determined, conclusions may be misleading.

The evidence that there is a range of intensities below the reinforcement value that may be detected comes from several directions. It was noted in our laboratory that when the intensity was switched to a low value, experienced rats and guinea pigs stopped responding after only one exposure to the stimulus. In contrast, if the stimulator was turned off, the animal typically responded 15 or more times before there was any interruption of a response rate. Evidently the low intensity stimulation was serving as a cue or signal for the beginning of a nonreward period. With no stimulation at all, the situation was more ambiguous. Recently, the quantitative relationship between the detection and reinforcement thresholds has been investigated by using intensities too low to maintain self-stimulation to signal the availability of reinforcing stimulus intensities (Campbell, 1963).

Stimulation intensities in the reinforcement range also have cue properties. This was demonstrated in an experiment in which positive stimulation served as a conditioned stimulus for an avoidance response (Mogenson & Morrison, 1962). Results with avoidance conditioning also suggest that electrical stimulation of any brain structure provides a distinctive cue to the animal regardless of any other effects associated with the stimulus (Nielson, Knight, & Porter, 1962).

The distinction between detection and reward threshold is equally important with aversive brain stimulation. In determining the threshold intensity for an aversive stimulus it was noted that the subjects (monkeys) were responding to the cue properties of the stimulus (Boren & Malis, 1961). In this testing procedure, an animal's response reduced the intensity of an aversive stimulus that was otherwise increasing at the rate of one step each second. The monkeys started responding when the stimulus reached a certain intensity and thereby prevented the stimulus from increasing above that point. It was first thought that this point indicated the aversive-stimulus threshold, but with additional experimentation it was recognized that as the intensity gradually increased, nonaversive stimuli were playing the role of signals which warned the animal of the forthcoming aversive stimuli. That animals were responding to the warning signal rather than to the aversive stimulation was shown by decreasing the maximal current. In this way it was possible to extinguish responding to the warning signals by separating them from the truly aversive stimuli.

It can be seen that any procedure which employs a regular order of presenting stimulus intensities may distort the reinforcement threshold. Estimates of the reinforcement threshold based on response rate are particularly likely to be distorted when the stimulus intensities are presented in a systematic pattern. With an ascending order, response rate may increase when the stimulus is detected because this serves as a signal for forthcoming reinforcement. This is similar to the well-established finding that with fixed intervals between reinforcements the response curve is "scalloped" (J function) due to the acceleration of rate as the reinforcement time approaches (Ferster & Skinner, 1957). Response rate in the range of intensities between the detection and reinforcement thresholds may reflect *anticipation* of reinforcement, but this is not equivalent to a *response* to a reinforcing stimulus. Anticipatory behavior may reflect general activity, as suggested above, and may also be influenced by a change in sensation threshold. As little is known concerning the mechanism by which animals detect the presence of the stimulus, it is evident that caution should be exercised before attributing a change in rate to a change in sensitivity of "reward" or "aversion" centers.

In contrast to a systematic sequence, a random sequence of intensity presentation yields data that are quite variable, as response rate produced by a given stimulus intensity is influenced by the intensity of the preceding stimulus. If the preceding intensity is high the response rate is likely to be significantly lower than that which would have been obtained with less intense preceding stimuli. Some preliminary data from our laboratory indicate that such effects may persist for as long as 30 minutes. It is not clear whether these results are due to an emotional reaction to the contrast in reinforcement strength or to some change in the responsiveness of the neural tissue following intense stimulation. We will raise this problem again, but at this point we would indicate that while a random presentation of stimulus intensities eliminates anticipatory behavior, it has the drawback of increasing variability.

Many of the problems discussed are exaggerated by a testing procedure which presents a series of intensities during a single testing session. The influence of both the cue properties of the stimulus and contrast effects due to differences in reinforcement strength are minimized when each session offers only one stimulus condition. Even under these conditions, however, it is difficult to make precise estimates of the reinforcement stimulus threshold from curves depicting rate as a function of stimulus intensity. The main problem is that criteria for distinguishing between sub- and suprathreshold performance are difficult to establish.

Figure 5 illustrates this difficulty and also compares results based on response rate with a time measure obtained from the two-platform test previously described. The similarity of the curves obtained with the two techniques supports the position that

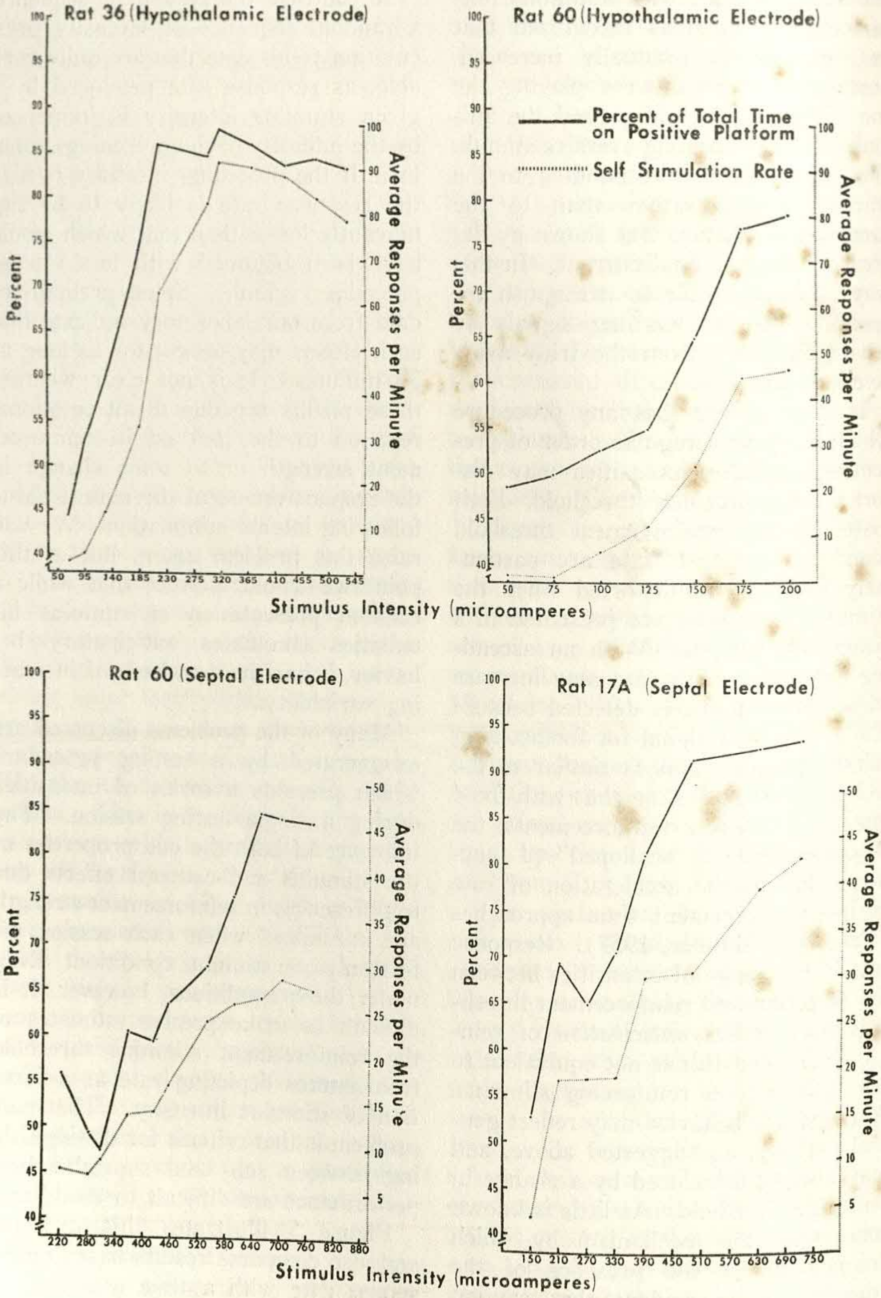


FIG. 5. Comparison of self-stimulation rate and percentage of total time on the positive platform with reinforcing hypothalamic and septal stimulation (after Valenstein & Meyers, 1964).

the same phenomenon was measured. However, it is instructive to compare the relative ease of determining the reinforcing stimulus threshold. With the two-platform test the positive (stimulation) and neutral platforms are switched on a random sequence throughout each test. As 50% of the total testing time on the positive platform represents chance expectancy, performance deviating from chance may be evaluated statistically. With self-stimulation rate there is no convenient standard against which to evaluate a specific rate. Response levels prior to stimulation experience (operant behavior) do not provide a useful criterion as it has been shown, for example, that even with one food-reinforced session, rates may consistently remain above previous operant levels (Segal, 1962). Also, after experience with reinforcing brain stimulation, we have noted that response rate remains significantly above "operant levels" during extinction trials.

Figure 6 illustrates that the response rate during the first two extinction tests averaged over 500% of the previous operant level. By the tenth test the rate had appeared to stabilize at approximately 200% of operant level. On the eleventh trial, the introduction of a 5-minute period of brain stimulation at the end of the 20-minute extinction session caused the rate to rise even in the nonstimulation period. Rate gradually declined with successive tests presumably as the animal learned that reinforcement was available only after 20 minutes. It would be anticipated that if reinforcing stimulation were introduced at random intervals, the rate during the

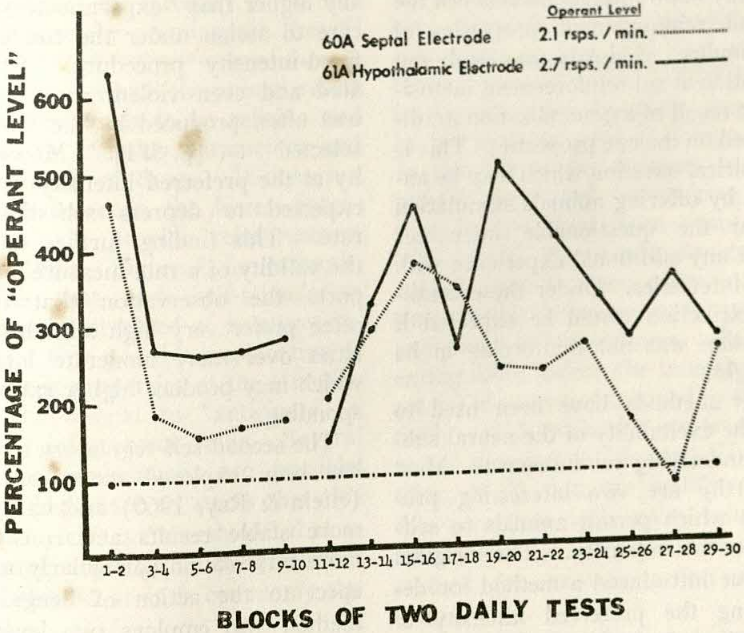


Fig. 6. Percentage of "operant" response level during extinction test following reinforcing brain stimulation. (Each point represents the average of two 20-minute daily tests. On Day 11 a 5-minute brain stimulation session was given following each 20-minute extinction period.)

extinction periods would have remained even higher.

This tendency toward higher response levels following a history of association with reinforcement illustrates the difficulty in estimating the reinforcement threshold precisely with a rate measure. It may be possible to obtain a daily operant level in order to compensate for this tendency toward increased response rates. For some purposes this method may be adequate, but in actual practice it will often be difficult (without an extensive investigation) to decide whether operant behavior should be sampled before, during, or after each test session.

After the animal has become familiar with the procedure, the reinforcement threshold obtained with the two-platform technique does not change with additional experience (Valenstein & Meyers, 1964), but there remains a possibility of an interaction between the cue and reinforcement properties of the stimulus. Animals may seek out stimulation at subreinforcement intensities as a result of a generalization gradient based on the cue properties. This is an empirical question which may be answered by offering animals stimulation only at the questionable intensities without any additional experience with higher intensities. Under these conditions extinction would be expected if stimulation was not reinforcing in its own right.

Other methods have been used to assay the excitability of the neural substrate underlying reinforcement. Most noteworthy are two interesting procedures which permit animals to self-regulate the intensity. The original technique introduced a method for determining the preferred intensity of stimulation (Stein & Ray, 1959). Animals were trained to press either of two levers to receive a brief reinforcing stimulus, but responses on one lever produced stepwise increments in intensity, while the other lever produced equal decrements in intensity. Current levels were started at zero and a trained animal increased the current level by responding on the appropriate lever and then by alternating between levers, maintained the intensity around a "preferred" level. This technique was apparently abandoned because many animals did not provide stable data. Current regulation was more reliable with posterior (hypothalamus and midbrain tegmentum) electrode placements, while rostral sites (septal area) produced poor regulation as animals often increased the intensity until convulsions occurred. In spite of this shortcoming, several interesting facts emerged. According to Stein and Ray (1959), preferred levels, for example, were usually higher than "experimenters would care to assign under the conventional fixed-intensity procedure. Exaggerated and even violent motor activity was often produced by the intensities selected . . . [p. 571]." Motor activity at the preferred intensity would be expected to depress self-stimulation rate. This finding further questions the validity of a rate measure and supports the observation that animals often prefer very high stimulus intensities over more moderate intensities which may produce higher rates of responding.

The second self-regulatory technique has been explored more extensively (Stein & Ray, 1960) and has yielded more stable results and considerably more information particularly with respect to the action of drugs. This method also employs two levers, but brain stimulation is received as a consequence of pressing only one of them. With each successive stimulus presen-

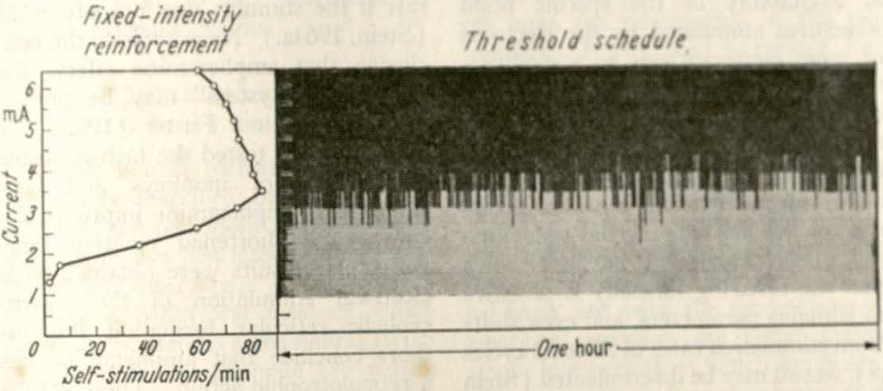


FIG. 7. Comparison of two indexes of the threshold for brain-stimulation reward: (left) self-stimulation rate obtained with fixed-intensity reinforcement versus (right) current intensity at reset in threshold schedule. (The horizontal markings to the left of the threshold record indicate the 16 current levels—15 steps—available from 6.4 milliamperes to zero milliamperes. In the 1-hour period shown, the animal gave himself 4,247 brain shocks and reset the current 558 times. After Stein and Ray, 1960.)

tation, the intensity is decreased in small steps, but at any time the animal may reset the intensity to the initial high level by pressing the "reset" lever.

The step at which the animal resets the intensity tends to be reliable and sensitive to stimulus parameters and administration of drugs. It does not represent, however, the lowest intensity that will provide reinforcement (Stein, 1961). Figure 7, for example, illustrates that high self-stimulation rates may be obtained at resetting intensities. One possible explanation is that the cue properties of the stimulus enable the animal to "anticipate" a drop in reinforcement value. Several alternative explanations are possible, but it is important to stress that this technique yields a "threshold" that is significantly higher than the reinforcement threshold as it is commonly defined.

There is one shortcoming of this procedure and one major question of interpretation which should be considered in view of the quantity of data derived from this approach. As with the self-regulatory method for preference threshold, this procedure does not appear to be useful with all reinforcing sites. To date, published studies include only electrode placements in the hypothalamus and tegmentum. In our laboratory attempts to obtain reliable data with reinforcing telencephalic sites have been unsuccessful. With septal placements, for example, animals tend to press more and more slowly with decreases in current level and often self-stimulation behavior is extinguished before the intensity is reset. This may be explained by the relatively rapid extinction reported with septal stimulation (Seward, Uyeda, & Olds, 1959), but the inability to use this procedure with many reinforcing areas limits its value particularly as a method for locating action sites of drugs.

Interpretively, changes in resetting level following administration of drugs have been attributed to either changes in excitability of the specific brain structures stimulated by the electrode or more conservatively to a modification of a positive reinforcing system. In order to understand any change brought about through the introduction of an experimental variable, it is important to determine what factors control the behavior in the test situation under analysis. It has been shown that the resetting intensity is sensitive to stimulus parameters, and even shifts from stimulation rates of 25 to 33 cycles per second may be discriminated (Stein & Ray, 1960). However, other variables in addition to the stimulus parameter-neutral excitability dimension may influence behavior in this situation.

In a test requiring repetitive acts, behavior generally develops a rhythmic pattern. In this particular test, for example, it may be observed that animals emit a regular number of responses on the stimulation lever, then press the reset lever and repeat this sequence over and over again. Any experimental variable that modifies this pattern will have an effect on the resetting intensity. Drugs such as amphetamine, pentobarbital, chlorpromazine, and reserpine which have been shown to alter resetting intensity have a number of both central and peripheral effects which may modify this behavior pattern other than through a presumed sensitization or depression of a reward system. The finding that amphetamine lowers the resetting intensity, for example, is viewed as resulting from facilitation of the hypothalamic reinforcing system (Stein, 1962b). The possibility that amphetamine induced a nonspecific motor activation was evaluated by testing animals in a conventional lever-pressing situation. Amphetamine increased response rates to a "subthreshold" stimulus, but there was no increase in rate if the stimulus was not presented (Stein, 1964a.) Nevertheless, the conclusion that amphetamine effects the "reinforcing system" may be premature. Uyeda and Fuster (1962), for example, have tested the tachistoscopic performance of monkeys and concluded that amphetamine improved accuracy and shortened reaction time. As similar results were obtained from electrical stimulation of the mesencephalic reticular formation the authors conclude that amphetamine has a reticulotrophic action. These results raise the possibility that it is the perception of subthreshold stimuli which may be influenced rather than the reinforcement process. In any case, a comparison of the conclusions from the two sets of experiments indicate the danger of attributing changes in performance to any specific or assumed neural system.

Amphetamine has been shown to increase the tendency to respond in many different experimental situations with both positive and negative reinforcement (see Carlton, 1963). In one illustrative experiment (Carlton, 1961) rats were required to alternate between two levers in order to obtain food reward. Following administration of amphetamine, "perseverative" tendencies were enhanced. That is, animals tended to repeat responses on the same lever although they received no food unless they alternated. With the self-determination method of obtaining brain-stimulation threshold animals were required to switch from one lever to the other in order to reset the stimulus intensity. Any "perseverative" tendency or enhancement of responding would cause the animal to drive the intensity down to a lower level. It would appear gratuitous at this time to attribute any change in resetting level to any alteration of specific neural elements.

NEURAL INTERACTION AND REINFORCING BRAIN STIMULATION

Limbic neural pathways have been extensively described, but the determination of the behavioral significance of these pathways lies outside the neuroanatomical discipline. It is necessary to have a behavioral measure which can be attributed to some neural structure in order to evaluate the nature of the interaction between structures. The self-stimulation phenomenon appears to be particularly convenient for this type of analysis as, at least as a first approximation, behavior may be considered to be under the control of the neural structures surrounding the electrode tip.

Through activation (stimulation) or deactivation (ablation or anesthetization) of one of two connecting structures, an indication of the interaction between these structures should be revealed. Using this basic approach some interesting interaction effects between specific limbic systems have been demonstrated, but the significance of the interaction is not always clear. For example, stimulation of the caudate nucleus of a monkey produced a high and stable rate of bar pressing at the start of a session. However, if stimulation of either the hypothalamus or amygdala preceded the caudate stimulation, lever pressing was significantly lower and less stable (Brady, 1961). To interpret such data we must ask whether these findings result from a specific interaction between these nuclei. With food reinforcement, for example, it has been observed that when an animal "anticipates" a particular reward a change in reward value will result in an elevation or depression of activity dependent upon the relationship of the two rewards. It has been suggested that this may result from an emotional reaction to the contrast (Crespi, 1944). It is not unreasonable to expect a similar effect with contrasting reinforcing brain stimulation, and indeed Hawkins and Pliskoff (1964) have reported such effects. Presumably contrast effects also have a neurological basis, but until the mechanisms of emotional reactions are better understood, interpretations which imply specific interactions between brain sites must be regarded cautiously.

Lesion technique may be used to determine whether a reinforcing site may be modified by eliminating the influence of another area. In one experiment the dependence of reinforcing tegmental stimulation on any crucial activity of the septal area or fornix system was tested (Ward, 1960). It was concluded that there was no such dependence, as rats with large lesions in the septal area continued to self-stimulate for tegmental stimuli. Providing the possibility of threshold change is examined, it would seem safe to draw conclusions from such "negative" results. However, where self-stimulation behavior is eliminated or decreased, it is necessary to determine whether the reinforcing properties of stimulation or some performance capacity has been modified. In a recent report it was shown that there were deficits in bar pressing for cingulate stimulation following hypothalamic lesions (Coons & Fonberg, 1963). The conclusion that the hypothalamus mediated reward obtained from cingulate stimulation must be evaluated together with the possibility that hypothalamic lesions affected performance sufficiently to cause a decrease in lever-pressing rate. As a minimum

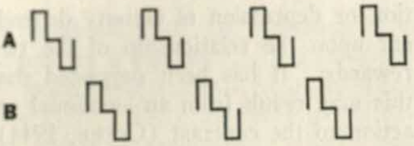


FIG. 8. Method of "simultaneously" stimulating two brain areas without electrical interaction (see text).

control it would be desirable to have a measure of the constancy of performance maintained by a reinforcer independent of brain stimulation.

Where results suggest an inhibitory action of one area on another it is also important to determine whether this action is specific or of a general nature. An example may serve to illustrate this point. There is a considerable body of information supporting the view that feeding behavior is controlled by the interaction of a lateral hypothalamic "feeding center" and a medial hypothalamic "satiety center" (Anand & Brobeck, 1951). As stimulation in the lateral area also provides positive reinforcement while stimulation in the medial area appears to be aversive, it has been suggested that there may exist a similar interaction of reinforcing systems (Hoebel & Teitelbaum, 1962). Indeed, it was demonstrated that when mild ventral stimulation was superimposed upon stimulation of the lateral area the animals stopped lever pressing. However, as medial stimulation is generally considered to have some aversive properties it is not surprising that the addition of a negative reinforcement should result in a lowering of response rate. It would be important to determine the influence of medial stimulation on other reinforcing brain sites (as well as with positive reinforcers other than brain stimulation) before any direct relationship is inferred.

Although studies of neural interaction that use electrical stimulation in contrast to lesion methods have the advantage of reversible effects, there are technical problems which raise some questions about interpretation. When knowledge has been obtained about the response of an animal to stimulation of a given neural area (Site A) there may be interest in changes in response characteristics resulting from stimulating a second neural area (Site B). It is hoped that some information pertinent to the nature of the neural interaction may be inferred by simultaneously stimulating the two areas. Not often considered is the possibility that *the interaction may be electrical rather than neural.* We have noted, for example, that effects characteristic of Site A (e.g., stereotype motor responses) may be triggered by stimulating Site B if a low, subthreshold stimulus is also presented to Site A. This may be true even when the stimulators are electrically isolated from one another and the electrodes located some distance apart. The fact that this effect may be produced when a number of different areas are substituted for Site B suggests that these results should not be attributed to a neural interaction. We do not at present understand the mechanism involved, but could suggest the possibility that overlapping electrical fields which are ineffective by themselves may interact in one of several ways to produce an observable result. In any case the problem may be circumvented. In our laboratory we are using an electronic gating circuit which rapidly opens and closes each electrode-stimulator pathway to prevent nonneural interactions. Figure 8 illustrates this method of "simultaneously" stimulating two sites. Both Stimulation A and B consist of biphasic pulse pairs presented at the rate of 100 per second, but each circuit is opened only during the presentation of a pulse pair. Other solutions are possible, but when simultaneous stimulation is used to study neural interaction, awareness of the possibilities of nonphysiological interactions should be indicated.

In addition to investigation of interaction between specific neural structures, other studies have dealt with the interaction between systems. Recently suggestions have appeared in the literature that an aversive-reinforcing system may inhibit the positive reinforcing system. In one experiment, rats placed in a shuttle box were permitted 7.5 seconds after the presentation of a tone to avoid aversive brain stimulation by crossing to the opposite side of the box (Stein, 1964b). Animals were slow to learn the avoidance response under these conditions and typically waited until receiving the first stimulation before leaping to escape. However, if a "priming" stimulation of a positive area was presented with the onset of the warning tone, the number of successful avoidance responses increased significantly.

How can these interesting results be interpreted? One way is to postulate an interaction between the positive and aversive neural systems. Such an interaction has been hypothesized by Olds and Olds (1962) with reference to positive lateral hypothalamic and aversive tegmental stimulation:

> the termination of the tegmental stimulus will result in a release phenomenon in the lateral hypothalamic area. That is, there will be an augmentation of neural activity in the lateral hypothalamus on termination of the supposed negative reinforcing stimulus. Now this augmentation may be the sole prerequisite of positive reinforcement [p. 809].

According to this theory, avoidance or escape behavior is not reinforced directly by the elimination of the noxious stimulation, but rather results from releasing the hypothalamic reinforcement center from an inhibitory influence. With respect to the performance of the rat in the shuttle-box situation described above, it has been suggested that the priming positive stimulation lowers the threshold in the hypothalamus and thereby increases the probability of the avoidance response (Stein, 1964b). Although the nature of the interaction is not explicitly stated, there is the implicit assumption that the reinforcement for the avoidance response is channeled through the hypothalamic system.

An alternative explanation emerges from an examination of the somatic concomitants of the stimulus. It can be observed that different motor patterns are brought into play by positive and negative stimulation. With positive brain stimulation animals move forward and appear to be actively investigating the environment. As the response is immediate and the stimulus provides no directional cues, it appears that this motoric response is directly triggered by the stimulation. Negative brain stimulation, provided it is not so intense that it causes the animal to jump straight up, appears to activate a "freezing" or backward movement. This triggering of competing responses (motor inhibition and facilitation) by stimulation of different limbic structures has been the subject of a number of studies (see Kaada, 1951). Behavioral implications of these findings have been studied by McCleary (1961). Where the environment is sensed by cephalad receptors, the evolutionary survival value of such behavior patterns would tend to result in their perpetuation. Noxious stimuli signaling danger elicit withdrawal patterns; positive stimuli elicit approach patterns.[4]

We are reminded of Guthrie's (1935) advice: "To train a dog to jump through a hoop, the effectiveness of punishment depends on where it is applied, front or rear [p. 160]." Recently, it was demonstrated that runway performance could be either facilitated or inhibited depending upon whether shock was delivered to the hind or forepaws (Fowler & Miller, 1963). It has already been noted that recent evidence has suggested that interoceptive as well as exteroceptive stimulation can directly determine response characteristics.

The motor patterns which are triggered by positive and negative stimulation are antagonistic. The warning tone (conditioned aversive stimulus) presented to the rat in the shuttle box elicits motor patterns which interfere with the forward movement necessary for successful avoidance. The "priming," positive stimulation initiates a forward movement, which improves avoidance performance in this situation. In agreement with this view is the observation that the positive stimulus antagonized the "freezing" reaction to the warning signal and permitted the avoidance response to get started (Stein, 1964b).

It is true that if the motor responses are antagonistic in a sense the underlying neural process may also be viewed as antagonistic. There is, however, an important distinction to be borne in mind. The antagonism of motoric reactions simply recognizes the fact that an animal can not be simultaneously immobilized and moving forward. This does not necessitate any direct interaction between central neural processes. Explanatory theories should take into consideration other alternatives as well as hypotheses that imply interactions within the central nervous system.

CONCLUDING REMARKS

The paper has dealt with problems of measuring the reinforcing consequences of brain stimulation. In illustrating problems there is always the danger of not viewing the progress in proper perspective. Actually much of the work cited has been of a pioneering nature, often extremely rich in hypotheses and frequently very creative methodologically. Subsequent work has attempted precision, but often with the rough implements of the frontier. Conclusions have gone beyond a general statement of possible usefulness of a technique to claims that this drug has this effect for this reason. It therefore seems appropriate and important at this time to examine the methods used to reach these conclusions.

Studies of the reinforcing consequences of stimulation are likely to make significant advances in our understanding of the physiology of such familiar chapter headings as motivation, emotion, drive, instinct, reinforcement, learning, and many others. It is most important that we travel a road which will permit the accumulation of valid conclusions which are meaningful beyond the idiosyncratic conditions of a particular experimental method. The present paper has not constructed any freeways. This has been an attempt to place caution signs where existing roads are dangerous and road blocks where progress is not possible. For the theoretician who attempts to integrate recent findings in this area:

PROCEED AT YOUR OWN RISK!

---

[4] For theoretical presentation of processes underlying approach and withdrawal mechanisms the reader is referred to Schneirla, 1959.

REFERENCES

ANAND, B. K., & BROBECK, J. R. Hypothalamic control of food intake. *Yale J. Biol. Med.*, 1951, **24**, 123–140.

ANAND, B. K., CHHINA, G. S., & SINGH, B. Effect of glucose on the activity of hypothalamic "feeding centers." *Science*, 1962, **138**, 597–598.

ANAND, B. K., DUA, S., & SINGH, B. Electrical activity of the hypothalamic "feeding centres" under the effect of changes in blood chemistry. *EEG clin. Neurophysiol.*, 1961, **13**, 54–59.

ASDOURIAN, D. Interaction effects of intracranial stimulation with rewarding and aversive solutions. *J. comp. physiol. Psychol.*, 1962, **55**, 685–690.

BOREN, J. J., & MALIS, J. L. Determining thresholds of aversive brain stimulation. *Amer. J. Physiol.*, 1961, **201**, 429–433.

BOWER, G. H., & MILLER, N. E. Rewarding and punishing effects from stimulating the same place in the rat's brain. *J. comp. physiol. Psychol.*, 1958, **51**, 669–674.

BRADY, J. V. Motivational-emotional factors and intracranial self-stimulation. In D. E. Sheer (Ed.), *Electrical stimulation of the brain.* Austin: Univer. Texas Press, 1961. Pp. 413–430.

BRADY, J. V., BOREN, J., CONRAD, D., & SIDMAN, M. The effect of food and water deprivation upon intracranial self-stimulation. *J. comp. physiol. Psychol.*, 1957, **50**, 134–137.

BRODIE, D. A., MORENO, O. M., MALIS, J. L., & BOREN, J. J. On the rewarding properties of intracranial stimulation. *Science*, 1960, **131**, 929–930.

CAMPBELL, J. A comparison of thresholds for the detection of subcortical electrical current and for self-stimulation. *Amer. Psychologist*, 1963, **18**, 436. (Abstract)

CARLTON, P. L. Some effects of scopolamine, atropine, and amphetamine in three behavioral situations. *Pharmacologist*, 1961, **3**, 60.

CARLTON, P. L. Cholinergic mechanisms in the control of behavior by the brain. *Psychol. Rev.*, 1963, **70**, 19–39.

COONS, E. E., & FONBERG, ELZBIETA. How hypothalamic and cingulate lesions affect the rate of rewarding self-stimulation in rats. Paper read at the Eastern Psychological Association, New York, April 1963.

CRESPI, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.

DELGADO, J. M. R., ROBERTS, W. W., & MILLER, N. E. Learning motivated by electrical stimulation of the brain. *Amer. J. Physiol.*, 1954, **179**, 587–593.

ECCLES, J. C. *The physiology of nerve cells.* Baltimore: Johns Hopkins, 1957.

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

FOWLER, H., & MILLER, N. E. Facilitation and inhibition of runway performance by hind- and forepaw shock of various intensities. *J. comp. physiol. Psychol.*, 1963, **56**, 801–805.

GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.

HAWKINS, T. D., & PLISKOFF, S. S. Brain-stimulation intensity, rate of self-stimulation, and reinforcement strength: An analysis through chaining. *J. exp. Anal. Behav.*, 1964, **7**, 285–288.

HODOS, W. Progressive ratio as a measure of reward strength. *Science*, 1961, **134**, 943–944.

HODOS, W. The nonaversive properties of long durations of rewarding brain stimulation. *Amer. Psychologist*, 1963, **18**, 437. (Abstract)

HODOS, W., & VALENSTEIN, E. S. Motivational variables affecting the rate of behavior maintained by intracranial stimulation. *J. comp. physiol. Psychol.*, 1960, **53**, 502–508.

HODOS, W., & VALENSTEIN, E. S. An evaluation of response rate as a measure of rewarding intracranial stimulation. *J. comp. physiol. Psychol.*, 1962, **55**, 80–84.

HOEBEL, B. G., & TEITELBAUM, P. Hypothalamic control of feeding and self-stimulation. *Science*, 1962, **135**, 375–377.

JUSTESEN, D. R., SHARP, J. C., & PORTER, P. B. Self-stimulation of the caudate nucleus by instrumentally naive cats. *J. comp. physiol. Psychol.*, 1963, **56**, 371–374.

KAADA, B. R. Somato-motor, autonomic and electrocortiographic responses to electrical stimulation of 'rhinencephalic' and other structures in primate, cat and dog. *Acta physiol. Scand.*, 1951, **24**, Suppl. 83.

MCCLEARY, R. A. Response specificity in the behavioral effects of limbic system lesions in the cat. *J. comp. physiol. Psychol.*, 1961, **54**, 605–613.

MILLER, N. E. Experiments on motivation. *Science*, 1957, **126**, 1271–1278.

MORGENSON, G. J., & MORRISON, MARJORIE J. Avoidance responses to "reward" stimulation of the brain. *J. comp. physiol. Psychol.*, 1962, **55**, 691–694.

NEWMAN, BERTHA L. Behavioral effects of electrical self-stimulation of the septal area and related structures in the rat. *J. comp. physiol. Psychol.*, 1961, **54**, 340–346.

NEWMAN, BERTHA L., & FELDMAN, S. M. Electrophysiological changes accompanying stimulation of the septal area in acute and chronic rat preparations. *Anat. Rec.*, 1960, **136**, 252.

NIELSON, H. C., DOTY, R. W., & RUTLEDGE, L. T. Motivational and perceptual aspects of subcortical stimulation in cats. *Amer. J. Physiol.*, 1958, **194**, 427–432.

NIELSON, H. C., KNIGHT, J. M., & PORTER, P. B. Subcortical conditioning, generalization and transfer. *J. comp. physiol. Psychol.*, 1962, **55**, 168–173.

OLDS, J. Runway and maze behavior controlled by basomedial forebrain stimulation in the rat. *J. comp. physiol. Psychol.*, 1956, **49**, 507–512.

OLDS, J. Self-stimulation of the brain. *Science*, 1958, **127**, 315–324.

OLDS, J. Hypothalamic substrates of reward. *Physiol. Rev.*, 1962, **42**, 554–604.

OLDS, M. E., & OLDS, J. Approach-escape interactions in rat brain. *Amer. J. Physiol.*, 1962, **203**, 803–810.

OLDS, M. E., & OLDS, J. Approach-avoidance analysis of rat diencephalon. *J. comp. Neurol.*, 1963, **120**, 259–295.

PLISKOFF, S. S., & HAWKINS, S. S. Some behavioral correlates of brain-stimulation reward. Paper presented at the Eastern Psychological Association, Philadelphia, April 1964.

PORTER, R. W., CONRAD, D., & BRADY, J. V. Some neural and behavioral correlates of electrical self-stimulation in the limbic system. *J. exp. Anal. Behav.*, 1959, **2**, 43–55.

REYNOLDS, R. W. The relationship between stimulation voltage and rate of hypothalamic self-stimulation in the rat. *J. comp. physiol. Psychol.*, 1958, **51**, 193–198.

SCHNEIRLA, T. C. An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1959*. Lincoln: Univer. Nebraska Press, 1959. Pp. 1–42.

SEGAL, EVALYN F. Prolonged extinction following one session of food-reinforced conditioning: A methodological note. *J. comp. physiol. Psychol.*, 1962, **55**, 40–43.

SEWARD, J. P., UYEDA, A., & OLDS, J. Resistance to extinction following cranial self-stimulation. *J. comp. physiol. Psychol.*, 1959, **52**, 294–299.

SIDMAN, M. *Tactics of scientific research.* New York: Basic Books, 1960.

SIDMAN, M., BRADY, J. V., BOREN, J. J., & CONRAD, D. G. Reward schedules and behavior maintained by intracranial self-stimulation. *Science*, 1955, 122, 830–831.

SPEAR, N. Comparison of the reinforcing effect of brain stimulation on skinner box, runway and maze performance. *J. comp. physiol. Psychol.*, 1962, **55**, 679–684.

STEIN, L. Inhibitory effects of phenothiazine compounds on self-stimulation of the brain. *Dis. nerv. Sys., Suppl.*, 1961, 22, 1–5.

STEIN, L. An analysis of stimulus-duration preference in self-stimulation of the brain. *J. comp. physiol. Psychol.*, 1962, **55**, 405–414. (a)

STEIN, L. Effects and interactions of imipramine, chlorpromazine, reserpine and amphetamine on self-stimulation: Possible neurophysiological basis of depression. In J. Wortis (Ed.), *Recent advances in biological psychiatry*. Vol. 4. New York: Plenum Press, 1962. Pp. 288–308. (b)

STEIN, L. New methods of evaluating stimulants and antidepressants. In J. H. Nodine & J. H. Moyer (Eds.), *Psychosomatic medicine, the first Hahnemann symposium.* Philadelphia: Lee & Febiger, 1962. Pp. 297–311. (c)

STEIN, L. Amphetamine and neural reward mechanisms. In H. Steinberg (Ed.), *Animal behavior and drug action*. London: J. & A. Churchill, 1964. (a)

STEIN, L. Reciprocal action of reward and punishment. In R. G. Heath (Ed.), *The role of pleasure in behavior*. New York: Hoeber, 1964. (b)

STEIN, L., & RAY, O. S. Self-regulation of brain-stimulating current intensity in the rat. *Science*, 1959, **130**, 570–572.

STEIN, L., & RAY, O. S. Brain stimulation reward "thresholds" self-determined in rat. *Psychopharmacologia*, 1960, **1**, 251–256.

UYEDA, A. A., & FUSTER, J. M. The effects of amphetamine on tachistoscopic performance in the monkey. *Psychopharmacologia*, 1962, **3**, 463–467.

VERHAVE, T. Towards an empirical calculus of reinforcement value. *J. exp. Anal. Behav.*, 1963, **6**, 525–536.

VALENSTEIN, E. S., & BEER, B. Reinforcing brain stimulation in competition with water reward and shock avoidance. *Science*, 1962, **137**, 1052–1054.

VALENSTEIN, E. S., & MEYERS, W. A rate-independent test of reinforcing consequences of brain stimulation. *J. comp. physiol. Psychol.,* 1964, 57, 52-60.

VALENSTEIN, E. S., & VALENSTEIN, THELMA. When does reinforcing brain stimulation become punishing? *Amer. Psychologist,* 1963, 18, 436. (Abstract)

WARD, H. P. Basal tegmental self-stimulation after septal ablation in rats. *Arch. Neurol.,* 1960, 3, 158-162.

WARDEN, J. C. *Animal motivation: Experimental studies on the albino rat.* New York: Columbia Univer. Press, 1931.

(Received October 16, 1963)

Psychological Review
1964, Vol. 71, No. 6, 438-456

ANALYZING THE COMPONENTS OF CLINICAL INFERENCE[1]

KENNETH R. HAMMOND, CAROLYN J. HURSCH, AND FREDERICK J. TODD

University of Colorado

The purpose of this paper is to analyze the components of clinical inference within the framework of Brunswik's lens model by means of multiple-regression analysis. 2 parallel studies of clinical psychologists, the performance of Ss in a quasi-clinical task, and the performance of Ss learning a multiple-cue probability task involving neutral stimuli provide the context for the analysis. Special reference is made to the problem of clinical vs. statistical prediction. Implications for the interrelation between experimental psychology, cognitive theory, and clinical tests are discussed.

Two sharply contrasting attitudes toward the psychology of inference may be observed within psychology. On the one hand, clinical psychologists behave as if there were virtually no limit to their capacity for inference making. Indeed, the general assumption, only slightly tarnished, is—the more data on which to base the inference, the better the inference is likely to be. On the other hand, experimental psychologists, particularly learning theorists, appear to believe that man's capacity for inference making is hardly more than a brief step removed from none at all. Where the truth lies, however, may be learned from the inferential behavior of the clinician himself, for clinical psychology provides an excellent research site for the study of man as an inferring organism.

This paper studies clinical inference within the framework of probabilistic functionalism, with Brunswik's lens model (1955, 1956) as the point of departure. The analysis is based on the application of multiple-regression technique to the lens model as initially set forth by Hammond (1955), and as recently outlined in further detail by Hursch, Hammond, and Hursch (1964).

LENS MODEL AND MULTIPLE-REGRESSION ANALYSIS

When the process of clinical inference is examined in terms of Brunswik's lens model (see Fig. 1), the application of multiple-regression analysis is straightforward. First, each half of the lens can be described in terms of the value of the multiple correlation coefficient (hereafter referred to as R_e for the environment, or stimulus conditions, and R_s for the subject). The R_e describes the relation between the cues and the variable to be estimated, and the R_s describes the relation between cues and the

---

[1] The research reported here was undertaken in the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and is Publication No. 39 of the Institute. This research was supported by a research grant (M-4977) from the National Institute of Mental Health. It is a part of the general program of research on inferential processes now in progress at the Behavior Research Laboratory.

The authors are indebted to Cameron R. Peterson, Jack L. Hursch, and Marilyn McRae for their assistance. We are especially grateful to Lawrence Grebstein, Joseph Newton, and Leon Rappoport for providing the authors with the original data from their studies.

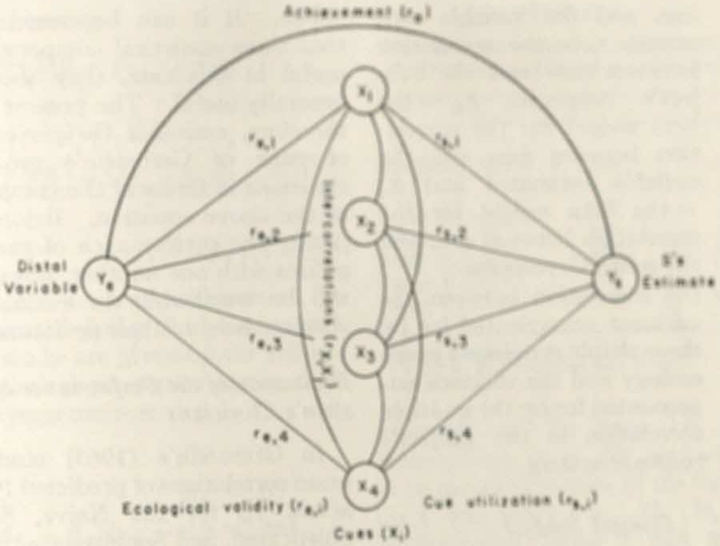


FIG. 1. Brunswik's lens model.

subject's responses. Thus, the degree of linear determinacy in each system can be calculated, and comparisons can be made between the environmental and organismic systems.

Second, the specific relation between each cue and the variable to be estimated (ecological validities) as well as the relation between the cues and the subject's response (utilization coefficients) can be measured in terms of correlation coefficients. Third, the sum of the differences between these correlations (more precisely, the sum of the product of the differences between ecological validities and utilization coefficients and the differences between the respective beta weights) can also be calculated. Such a calculation provides an index of the extent to which the subject uses each cue relative to the validity of the cue.

Fourth, a partial correlation between the subject's judgment and the variable to be estimated (with the linear variance of each system eliminated) may be calculated.

The relation between overall achievement (correlation between the subject's estimate and the variable estimated) and the statistical components described above is set forth in the equation below (see Hursch et al., 1964, for a full development of the equation):

$$r_a = \frac{R_e^2 + R_s^2 - \Sigma d}{2} + C\sqrt{(1 - R_e^2)(1 - R_s^2)}$$

where:

r_a = the correlation between subject's judgments and the variable estimated

R_e = the multiple correlation between the cues and the variable estimated

R_s = the multiple correlation between the cues and the subject's judgments

Σd = the sum of the products $(r_{e_i} - r_{s_i})(\beta_{e_i} - \beta_{s_i})$ where r_{e_i} = the correlation between

cue_i and the variable estimated, r_{s_i} = the correlation between cue_i and the subject's judgment, β_{e_i} = the beta weight for the correlation between cue_i and the variable estimated and β_{s_i} = the beta weight for the correlation between cue_i and the subject's response

C = the correlation between the variance unaccounted for by the multiple correlation in the ecology and the variance unaccounted for by the multiple correlation in the subject's response system

Clinical Study I

Within the course of comparing the predictive efficiency of the multiple regression equation with that of clinical judgment, Grebstein (1963) compared the performance of five *Naive* clinicians (graduate students who had completed an introductory course in Rorschach administration and scoring), five *Semisophisticated* clinicians (graduate students who had completed the above training and were also second-year Veterans Administration trainees), and five *Sophisticated* clinicians (professional clinical psychologists with the PhD and 5 years of experience) whose task was to estimate 30 patients' (Wechsler-Bellevue) IQs using 10 cues from their Rorschach (Beck) psychograms. Grebstein (1963) concluded that "ability to judge IQ from Rorschach psychograms does not improve with additional clinical experience [p. 130]."

Irrespective of the substantive nature of Grebstein's results, his study provides an excellent site for the examination of the heuristic value of the statistical components discussed above. If it can be demonstrated that these statistical components are useful in this case, they should be generally useful. The present study, therefore, examines the performance of each of Grebstein's groups of clinicians in terms of the components of the above equation. Before comparing the performance of the three groups with one another, however, it will be worthwhile to evaluate the absolute level of their performance.

Evaluation of the Performance of Grebstein's Clinicians

In Grebstein's (1963) study the mean correlations of predicted IQ with actual IQ for the Naive, Semisophisticated, and Sophisticated Judges were .50, .65, and .68, respectively. While these correlations were found to be significantly different from zero, they fall far short of 1.00. Thus Grebstein concludes, "Experienced clinical psychologists can make accurate estimates of intelligence from Rorschach summaries, but there is room for improvement [p. 132]." The question arises, however, as to *how much* room for improvement exists. The following section will demonstrate that the limits of achievement will depend upon the relation between characteristics of the test situation and the characteristics of the response system of the clinician.

Limits of Achievement. The principal aim of Hursch et al. (1964) was to examine six different relations between the statistical properties of the environment, or task, and the statistical properties of the response system in order to discover the limits of achievement in each case under various conditions. Grebstein's study falls under Case VI, in which it is shown that achievement (r_a), the correlation between the clinician's estimate of IQ and the patient's IQ, must lie between:

$$\frac{R_e^2+R_s^2-\Sigma(\beta_{e_i}-\beta_{s_i})(r_{e_i}-r_{s_i})}{2}-C\sqrt{(1-R_e^2)(1-R_s^2)}\leq r_a$$
$$\leq\frac{R_e^2+R_s^2-\Sigma(\beta_{e_i}-\beta_{s_i})(r_{e_i}-r_{s_i})}{2}+C\sqrt{(1-R_e^2)(1-R_s^2)}$$

Thus, the limits of achievement for Sophisticated Judge 5 in Grebstein's (1963) study are given by the following statistical properties of the task and his response system:

$$R_e^2 = .63$$
$$R_s^2 = .85$$
$$\Sigma(\beta_{e_i}-\beta_{s_i})(r_{e_i}-r_{s_i}) = .264$$

and since the limits of C are ±1, then the limits of r_a are:

$$\frac{.63+.85-.264}{2}-\sqrt{(.37)(.15)}\leq r_a\leq\frac{.63+.85-.264}{2}+\sqrt{(.37)(.15)}$$
$$.608-.234\leq r_a\leq .608+.234$$
$$.374\leq r_a\leq .842$$

Therefore, Sophisticated Judge 5 whose achievement, r_a, was actually .68, would have reduced his achievement to .37 if the residual variance of his response system had been perfectly inversely correlated with the residual aspects of the Rorschach-IQ system, and would have reached .84, had this correlation been positive and unity.

If the correlation between the two residual variances had been exactly zero, then:

$$r_a=\frac{R_e^2+R_s^2-\Sigma d}{2}$$
$$=\frac{.63+.85-.264}{2}$$
$$= .61$$

which is not far from his actual achievement, (r_a = .68), and which is to be expected in view of the fact that his C was fairly low, .32. If, then, Sophisticated Judge 5 was not expected to develop a substantial value of C (and such an expectation implies a great deal), his achievement was very close to the maximum he could have achieved under the circumstances. This point will be developed further as each component is discussed in detail. What is to be remembered is that the upper limit of achievement is not *uncritically* to be assumed to be unity.

TABLE 1

MULTIPLE-REGRESSION COEFFICIENTS FOR INDIVIDUAL JUDGES AND GROUPS[a]

	Naive	Semisophisticated	Sophisticated
Median	.905	.951	.919
Judge 1	.952	.898	.863
Judge 2	.871	.951	.928
Judge 3	.894	.955	.930
Judge 4	.979	.852	.870
Judge 5	.905	.960	.919

Note.—Kruskal-Wallis H = .98; p = ns.
[a] Based on data from Grebstein (1963).

Differences between Naive, Semisophisticated, and Sophisticated Clinicians

Tables 1, 2, and 3 show (*a*) that clinicians with varying amounts of training and experience did not differ with respect to R_s, the linearity of the response process, (*b*) that they did not differ with respect to C, the correlation between the residual variance in the Rorschach-IQ system and residual variance in the subject's response system, but (*c*) that they did differ with respect to Σd, the degree of match between ecological validities and utilization coefficients. Each of these results is discussed in turn below.

Linearity (*R*)

The fact that the three groups did not differ with respect to R_s (see Table 1) indicates that they did not differ in terms of the linearity of their response systems. The significant fact for the present discussion is that all groups are highly linear. Moreover, they are more linear than the Rorschach-IQ system of relations ($R_e = .79$). The median R for the Naive clinicians is .905; for the Semisophisticated clinicians, the median $R = .951$; and for the Sophisticated clinicians, the median $R = .919$. The rank order correlation between linearity (R) and achievement (r_a) is negligible.

The question of linearity is worth discussing in detail because it aids in clarifying the issues involved in statistical and clinical prediction. Note that if the clinicians were to increase their linearity to $R_s = 1.00$ in such a way that their judgments fell on the regression plane defined by the multiple correlation characterizing the relationship between Rorschach cues and IQ, their achievement correlation would be just equal to R_e, or .79. (See Case III/1, Formula 17, in Hursch et al., 1964.) The mean achievement ($r = .68$) of the Sophisticated clinicians as computed by Grebstein (1963, p. 129), was quite close to this, as *might* be expected from the high linearity of the clinicians. On the other hand, the multiple-regression technique, in which the correlation between the cues and the estimate is exactly 1.00, produced an achievement r_a of only .56 on a *cross-validation sample*. The difference between the achievement of the clinicians and the multiple-regression technique raises an interesting point which has received considerable partisan discussion (Holt, 1958; Meehl, 1954; Sarbin, Taft, & Bailey, 1960). Each procedure is discussed below in terms of the linear process.

Linearity and Clinical Prediction. Because the clinician's linearity is extremely high, about .90, for the moment we shall ignore this slight departure from unity and assume that they were indeed perfectly linear ($R_s = 1.00$). And because the achievement ($r_a = .68$) of the composite Semisophisticated and Sophisticated clinicians sufficiently approximates the maximum attainable (.79) if they had been perfectly linear, it is also plausible to assume that the clinicians achieved the maximum accuracy an R_s of 1.00 and an R_e of .79 would permit. If the clinicians had actually performed according to these assumptions (which are close to the fact), how would such performance be explained?

If R_s is taken to be 1.00, then it would have to be concluded that a linear, additive *rule* was being used. What would such a rule consist of? One answer is that the clinicians were either systematically applying an *inductive* rule based on experience, or were systematically applying a rule taken from a *textbook*. There is some evidence that the Sophisticated clinicians were doing the former, and the Naive clinicians the latter. The evidence comes from an inquiry into the specific mismatches between ecological validities and utilization coefficients in Grebstein's study.

When the differences between the utilization coefficients and ecological validities of the three groups are examined with respect to the 10 Rorschach factors under consideration, it becomes apparent that the Naive clinicians misused one cue, Beck's Z[2], far more than other cues. The ecological correlation between Z and IQ in Grebstein's sample is .18. The mean utilization coefficient is .63 (corrected) for Z in the Naive group, however, which is much higher, whereas the mean utilization coefficient for the Semisophisticated groups is only .38 and .27 for the Sophisticated clinicians, indicating much closer approximation to the empirical value of .18. It appears plausible, then, that the Naive clinicians were relying on a textbook rule, whereas the Sophisticated clinicians were depending on a rule developed from experience.[3]

---

[2] Beck (1946) described the relation between Z and intelligence as follows: "it (Z) is an index of the intellectual energy as such, irrespective of the *kind* of intelligence that

The highly accurate predictions made by the experienced clinicians would have to be explained in terms of optimal weighting. Given a multiple R_s of 1.00, there is only *one* set of weights which could be applied to the regression equation which would result in maximum achievement. The high achievement indicates that the experienced clinicians were actually approximating this unique set of weights very closely.

Although it might be argued that the clinicians were aware of the rule they were applying, and were consciously applying it with precision so that the R_s closely approached 1.00, it strains credulity to believe that they were also consciously applying optimal weights with the same precision so that achievement approached a maximum.[4] Because these conditions were very nearly met, however, the results present a challenge to cognitive theory.

Linearity and Prediction by the Multiple-Regression Equation. The performance of the multiple-regression equation, it will be recalled, provided an achievement correlation of .56. In Grebstein's procedure the multiple-regression equation was developed on the basis of one sample and tested on a second, cross-validation sample. It is important to note that when the multiple-regression equation is applied to the cross-validation sample, the "response" of the multiple-regression equation is completely determined by the equation derived from the initial sample. In other words, $R_s = 1.00$. Whatever pecularities existed in the first sample now are applied by the equation to the second in a deterministic way, precisely because $R_s = 1.00$. Whatever differences exist between Sample 1 and the cross-validation sample, then, may be expected to *decrease* the achievement correlation between the estimates provided by the equation and the actual IQs to a level below the maximum of .79; in this case r_a was reduced to .56.

The above point, which concerns shrinkage, has not always been observed. In defending clinical prediction against actuarial prediction, Holt (1958), has argued that it is unfair to pit empirically adjusted equations against clinical predictions which have not had the benefit of such empirical correction. Holt then proceeds to compare clinical predictions with a multiple-regression equation developed from *one* sample, which was in no sense random, and finds that the

---

S uses, something that does influence W. Thus, Z is a more accurate representative of the intelligence functioning per se. . . .

"These higher Z scores are always found in individuals whose work requires most difficult and complex intellectual effort. Like W, it is therefore an index of thinking power. Its essence is the capacity to grasp relations not perceived by others [p. 12]."

[3] See Hammond, Kern, Crow, Githens, Groesbeck, Gyr, and Saunders (pp. 526–542, 1959) for a detailed examination of differential dependence on cues as a function of professional interest on the part of raters.

[4] Both Todd (1954) and Hoffman (1960) found that their subjects could not report accurately on their use of cues. See also Rommetveit (1960).

TABLE 2

Correlation (C) between Nonlinear Aspects of Judges' Response Systems and Rorschach-IQ Systems for Individual Judges and Groups[a]

	Naive	Semisophisticated	Sophisticated
Median	.229	.179	.144
Judge 1	−.057	.328	.144
Judge 2	.229	.035	.107
Judge 3	.244	.179	.322
Judge 4	.039	.214	−.056
Judge 5	.390	.051	.320

Note.—Kruskal-Wallis $H = .00$; $p =$ ns.

[a] Based on data from Grebstein (1963).

predictive utility of the equation disappears when cross-validated on a new and, of course, *different* "sample" (cf., Sarbin, Taft, & Bailey, 1960, pp. 247–250). If the samples are not random, the multiple-regression equation is not only subject to shrinkage due to sampling errors, it is subject to systematic error as well. (If samples are indeed random, then, of course, the shrinkage of the multiple R can be computed directly (Guilford, 1956; McNemar, 1962.)

On the other hand, neither Grebstein's nor Holt's clinicians were required to have their performance completely determined by the idiosyncracies of a prior sample. In effect then, the multiple-regression technique is put to an unfair test whenever it is calculated from a single, small sample, because its "experience" is thereby severely limited. If the equation in Grebstein's study had been allowed to develop from greater experience, however, either as a result of exposure to several random samples, or through exposure to a very large N, it would in all likelihood have performed much better.

Ironically, the highly linear performance and higher r_a of the clinician lends support to this argument. For there was a remarkable similarity between clinical inference and the multiple regression model in this study. The clinicians' inferential processes were nearly identical with the multiple-regression procedure both in *function and in content* (see Grebstein, p. 129). Both procedures used essentially the same variables (M, F + %, Z, and RC) in the same, i.e., linear, way. The clinicians therefore, were proving how good an "experienced" linear process could be. If there is a dispute between clinical and statistical prediction, it would be difficult in this case to discriminate between the contestants' performance.

Since the purpose of this paper is heuristic, rather than substantive, let us now assume that the clinicians' responses were *not* completely identical with a multiple-regression equation based on any sample no matter what size. Indeed, there is a sense in which such linear indeterminacy defines the clinical method. This lack of complete linear determinacy in the clinicians' response system suggests the possibility that they utilize nonlinear as well as linear relations in making their judgments—a point to which we now turn.

Correlation between Residual Variances (C)

The data from Table 2 indicate that there were no differences between the three groups of clinicians in this respect.

The analysis of C in Grebstein's study has direct implications for test construction. First note that if $R_s = 1.00$, there is, of course, no residual variance in the clinician's response system, and therefore no correlation is possible with the residual variance in the environment. Whenever $R_s < 1.00$, however, the possibility exists that the clinician can raise his achievement by correlating the residual variance in his response system with that in the environment, provided R_e is also <1.00. (This possibility is not available to the multiple-regression equation because residual variance is treated only as random error.) The important point to be observed here is that C can become large only if there is some systematic, nonlinear variance in the environmental system which can be detected by the clinician. Because the multiple-regression procedure *cannot* improve its performance in this way, *the magnitude of* C *specifies the extent to which the clinician effectively uses those special properties which distinguish him from the multiple-regression equation.* Of course, if the task does not provide the opportunity, he cannot use them.

Since all of Grebstein's clinicians had R_s's slightly less than unity (ranging between .84 and .98) and since approximately 40% of the variance in the Rorschach-IQ system is not accounted for by linear regression ($R_e^2 = .63$), the question arises as to the extent of the correlation between the nonlinear variances in the two systems. The values of C in Table 2 are quite low, ranging from $-.06$ to $+.39$, indicating little, if any, relation between the residual variances. Whether the residual variance in the Rorschach-IQ system is to be considered as error, or as systematic, *non*linear variance remains an open question. The obtained values of C indicate, however, that if any systematic nonlinear variance existed, it was not, in the main, detected and correctly utilized by the clinicians in Grebstein's study.

This result raises several questions. Does the Rorschach-IQ system in fact have systematic, nonlinear variance which is detectable? Should not this question be answered in a definitive form before requiring clinicians to attempt to detect such relations? A further question is: given the demonstrated *linear* relations between the Rorschach cues and IQ, would systematic, detectable nonlinear variance make an appreciable difference in achievement even if it were detected?

The answer to the latter question can be found by referring to the limits of achievement calculated earlier for Sophisticated Judge 5. There we observed that had this Judge maintained the same degree of linearity ($R_s^2 = .85$) and the same degree of match ($\Sigma d = .264$), but increased C from .32 to 1.00, his achievement would have increased from .68 to .84. In view of the highly remote likelihood of achieving such a perfect correlation, the effort first, to *discover* perfect nonlinear relations in the Rorschach-IQ system, and second, to *train* someone to achieve perfect recognition of these relations without reducing an already creditable overall achievement makes such an investment appear to be a very poor risk indeed. Furthermore, if the clinician did achieve a C of 1.00, he would also have to obtain an R_s^2 exactly equal to R_e^2 (.63), and a Σd of exactly zero in order to obtain perfect achievement ($r_a = 1.00$). Perfect matching of R_e^2 and Σd on the part of the clinician seems much less likely than perfect (or near perfect) linearity.

These conclusions bring us to the fundamental implications of the above considerations for test construction itself.

Nonlinearity and Test Construction. Those who develop psychological tests which require clinicians to make inferences from the nonlinear aspects of the cue data provided by the tests should bear the burden of demonstrating that such nonlinear relations (*a*) exist, (*b*) are detectable by clinicians, and (*c*) are worth whatever trouble it takes to detect them. Moreover, test makers will have to take into consideration the fact that our scientific knowledge concerning the human's ability to detect nonlinear relations is hardly more than primitive. Indeed, the entire question of the nature of human information processing is still in its infancy. In view of all this, one wonders whether the present methodology for construction of tests which require clinical inference has not outrun our knowledge of cognitive processes, and whether future makers of such tests should not include in their methodology considerations of the psychology of inference (the "cognitive activity of the clinician" as Meehl, 1960, has put it) as well as the psychology of personality. This calls for nothing more than accepting the operational consequences of what everyone has long accepted—the notion that both test and clinician are interacting parts of the same system (cf. Hammond, 1955).

This discussion does not, of course, exhaust the implications of nonlinearity for test construction, but that topic is too large to be explored fully here. We turn now to a consideration of the third component of the analysis, Σ*d*.

TABLE 3

ΣD FOR INDIVIDUAL JUDGES AND GROUPS[a]

	Naive	Semisophisticated	Sophisticated
Median	.700	.458	.331
Judge 1	1.903	.588	.331
Judge 2	.834	.478	.421
Judge 3	.652	.458	.509
Judge 4	.367	.276	.180
Judge 5	.700	.289	.264

Note.—Kruskal-Wallis $H = 6.5$; $p < .05$.
Naive versus Semisophisticated $p = .028$.
Naive versus Sophisticated $p = .016$.
Semisophisticated versus Sophisticated $p = .210$.

[a] Based on data from Grebstein (1963).

TABLE 4

DIFFERENCES IN r_a AMONG THREE GROUPS OF CLINICIANS VARYING IN EXPERIENCE[a]

Mann-Whitney *U*	
Naive versus Semisophisticated	$p = .028$
Naive versus Sophisticated	$.008 < p < .016$
Semisophisticated versus Sophisticated	NS
Kruskal-Wallis $H = 7.215$	$.02 < p < .05$

[a] Based on data from Grebstein (1963).

The Degree of Match between Ecological Validities and Cue Utilization (Σ*d*)

Grebstein's three groups do not differ with respect to (*a*) the linearity of their response systems (*R*), or (*b*) the extent to which they were able to detect systematic, *nonlinear* combinations in the residual variance of the Rorschach-IQ system (*C*); the three groups *do* differ with respect to (*c*) the extent to which utilization coefficients match ecological validities, however.

The data in Table 3 show that there were statistically significant differences between Grebstein's groups with respect to Σ*d*; this component becomes progressively smaller, as it should, as the experience of the clinician increases. Furthermore, Σ*d* is highly related to *achievement* within groups; the smaller Σ*d*, the higher the achievement. (Rank difference correlations for the Naive, Semisophisticated, and Sophisticated groups are −.90, −.90, and −.72, respectively.) Σ*d*, therefore, clearly has promise as an important component of clinical inference.

The fact that the three groups differed with respect to Σd, and that the within-group correlations between Σd and achievement were quite large, led to a re-examination of Grebstein's conclusions that the three groups did not differ with respect to achievement. When the achievement correlations in each group of clinicians were analyzed in terms of the Mann-Whitney *U* test (the results of Bartlett's Test of homogeneity having indicated a parametric test to be inappropriate), significant differences between groups were obtained. The results of this analysis (see Table 4) are consistant with the analysis of Σd—progressive differences in achievement are observed as experience increases, although the difference between the Semisophisticated and the Sophisticated clinicians is not statistically significant.

The inference that achievement might well be increased by increasing the match between ecological validities and utilization coefficients, is therefore supported by the data in Table 4 which indicate that the groups with smaller Σd are the groups with greater achievement. (As pointed out above, the specific mismatch which led to a large Σd for the Naive clinicians was the overdependence on Beck's Z.) Σd, however, is a highly complicated concept, and its relation to the use and misuse of cues is not unambiguous.

Σd and the Weighting of Cues. The quantity Σd, while serving as an important component of achievement, does not easily lend itself to further analysis for two reasons. The first reason is the highly interdependent relations of the components of the multiple-regression analysis presented here. Changes in the value of any one of these components can result in changes in the values of other components. The value of Σd itself as well as its relation to r_a is dependent on the values of the other components. The contribution of Σd to r_a, therefore, cannot be assessed without examination of the other components as well. Thus, our treatment of Σd, in conjunction with our consideration of R_e^2 and R_s^2, with an assumed or computed value for C, is a valid one, whereas the consideration of Σd alone might not be.

Secondly, when Σd is computed from empirical data as in the present analysis, one may be tempted to look at each difference product, compare these with each other, and make statements about the relative contribution of each cue to Σd according to the size of its difference product. This, however, would be a hazardous procedure because the quantity Σd is not a simple sum of the mismatches, rather it is a resultant of interdependent difference products (or vectors). For this reason, it is impossible to evaluate the independent effect of any one mismatch on the resultant Σd. (cf. Fink, 1961; Hoffman, 1960; Ward, 1962.)

Matching versus Linearity

Earlier we discussed the implications of perfect linearity on the part of the clinician. There we found that achievement, r_a, could be as large as R_e. Under conditions of *perfect matching* of utilization coefficients with ecological validites (more precisely, conditions under which Σd equals zero), r_a equals R_e^2, if we assume $C = 0$ (see Case II, Equation 13, in Hursch et al., 1963).

Thus, if one of Grebstein's clinicians had a Σd which equalled zero (and C also $= 0$), his achievement, r_a, would be exactly equal to the *square* of the multiple correlation in the environment, .63. Since $R^2 < R$, perfect matching ($\Sigma d = 0$) always results in a *lower* limit of achievement than the limit provided by perfect linearity (assuming $C = 0$). This is a point worth remembering when comparing clinicians with the linear multiple-regression equation.

Consider now Semisophisticated Judge 4 who had a high achievement correlation of .60, a very small Σd (.275) and a relatively low R (.852). Consider also Naive Judge 1 who had the lowest achievement correlation (−.196), a very large Σd (1.90), and a very high R (.952). According to our earlier remarks, we conclude that because Semisophisticated Judge 4 has a smaller R, he is less rule bound than Naive Judge 1. Judge 4, however, has a higher r_a in spite of this because his match is better, that is, his Σd is smaller. Should Naive Judge 1 become less rule bound or should he try to decrease Σd? The choice between changing R or Σd bears implications for cognitive theory related to the above discussion concerning linearity and the clinician.

Matching and Cognitive Theory

Earlier it was asserted that increasing linearity implied increasing rule-bound behavior. Decreasing such behavior, then, implies moving in the direction of probabilistic behavior.

Brunswik's (1956, p. 89–99) theoretical distinction between thinking and perception parallels the above distinction between rule-bound and probabilistic behavior. According to Brunswik, at one pole of the cognitive continuum lies "certainty-geared" deductive thinking, and at the other pole lies "uncertainty-geared" intuitive perception. Deductive thinking implies rule-bound cognition. And probabilistic, intuitive perception should involve the *matching* of ecological validities of perceptual cues with their respective utilization coefficients (1956, p. 141). Between these two poles lies *quasi-rationality* (Brunswik, 1954; Hammond, 1961; Heider, 1958) an area in which both types of cognition participate. How a subject distributes or employs these two types characterizes his inferential mode. The effectiveness of his inferential mode is determined by the properties of the environment.

Within the Brunswikian framework, then, to urge Naive Judge 1 to give up his rule-bound response system and to increase his match implies that he is to be urged to give up an analytical approach and to adopt an intuitive one. And to urge Semisophisticated Judge 4 to increase the rule-like quality of his response system without decreasing his match is to urge him to become more analytical *without* giving up his successful intuitive performance. Our knowledge of cognition, of course, does not permit us more than conjectures as to whether such changes are within the realm of possibility. A study by Rappoport (1963) however, indicates that pairs of subjects learning under an intuitive set engage in less interpersonal conflict than pairs of subjects learning under an analytical set. This result suggests that the response systems of rule-bound subjects are harder to change than those of intuitive subjects, and implies that it would be easier to change the more intuitive clinician (here Semisophisticated Judge 4) than the more analytical one (Naive Judge 1).

We turn now to a brief examination of the performance of 10 experienced clinicians with respect to all three components of performance.

Clinical Study II

An examination of Todd's (1954) 10 experienced clinicians shows that these clinicians (*a*) were also highly linear in their response systems (median R of .93),[5] (*b*) had a mean Σd about the same as Grebstein's (.044 versus .055) and, (*c*) had C values approximately equal to those of Grebstein's clinicians, although seven of Todd's clinicians had ***negative*** Cs, which reduced their achievement (see Hursch et al., 1964, Table 5, for complete details). As in the case of Grebstein's clinicians, the rank order correlation between Σd and achievement was substantial, being equal to −.76, while the rank-order correlation between R_s and r_a was negligible (.22).

In brief, Grebstein's results confirm those obtained by Todd. One factor contributing to this similarity of results may be that the multiple correlation between Rorschach factors (Klopfer technique) and IQ in Todd's study was .77, and in Grebstein's study (Beck technique) was .79.

A QUASI-CLINICAL STUDY

Newton (1963) carried out what might be called a study of quasi-clinical inference. His subjects were 99 college sophomores who estimated Freshman grade averages ($N = 53$) from four cues (IQ, high school rank, a College Board Score, and a personality rating by the high school principal). These subjects were also highly linear (mean R_s over five conditions approximately .90), had a mean Σd identical with Todd's clinicians ($\Sigma d = .05$) and had negligible values of C.

Newton's subjects were first asked to make their estimates of grade average on the basis of their campus experience. Four different types of feedback were then employed, each type providing progressively more information. (The fifth condition required the subjects to form an explicit inference rule.) The data of Table 5 show the rank-order correlations between Σd and achievement, r_a, in each condition. Although these correlations are not as high as those obtained with the clinical psychologists described above, they indicate substantially the same trend, a definite relation between Σd and r_a.

TABLE 5

RANK ORDER CORRELATIONS BETWEEN ΣD AND r_a BEFORE (B) AND AFTER (A) DIFFERENT CONDITIONS OF FEEDBACK*

Conditions									
I		II		III		IV		V	
B	A	B	A	B	A	B	A	B	A
.36	.41	.37	.37	.69	.71	.55	.65	.34	.46
$n = 21$		$n = 20$		$n = 19$		$n = 20$		$n = 19$	

* Data from Newton (1963).

Implications of Variation in Feedback for Improving Achievement

Newton's study illustrates a further point concerning clinical inference discussed by Holt (1958, p. 5). In disputing the current methodology employed in comparing the efficiency of clinical and actuarial methods of prediction, Holt argued that the clinician's performance should be exposed to the same sort of self-correction employed in actuarial methods. That is, just as the statistician examines the validity (or weights) of various items or tests, the clinician should be allowed to examine the empirical validity of the variables he is depending upon for his inferences.

Although Newton's study was carried out for different purposes, his results bear directly on Holt's point. His subjects estimated Freshman grade averages on the basis of four cues, as indicated above. Each subject was then assigned to one of five different conditions in which different forms of feedback were provided (see Fig. 2). The results indicate that those subjects who were told what the ecological validities in the sample were, and what their utilization coefficients were, improved (not substantially, but to a statistically significant degree) in their second set of 53 trials.

---

[5] The high R_s's obtained with Grebstein's clinicians are of course, subject to shrinkage. The high R_s's obtained in Todd's study, however, were maintained on cross validation.

The fact that college sophomores improved their achievement after *one exposure* to the ecological validities and to their own manner of utilizing cues (and a brief explanation of the concept of correlation) seems quite remarkable. Whether it opens the door to new training methods for clinicians—with respect to certain inferential tasks, at least—remains to be seen. Holt's position, in any event, is confirmed by Newton's results. If subjects are provided with the appropriate information concerning cues and their responses, they can improve their performance. Lee and Tucker (1962) report a similar result with respect to a multiple-probability learning experiment.[6]

A MULTIPLE-CUE PROBABILITY LEARNING EXPERIMENT

As a part of a larger study, Rapport asked subjects to estimate the values of numbers on the back of 3 × 5 inch cards on the basis of the

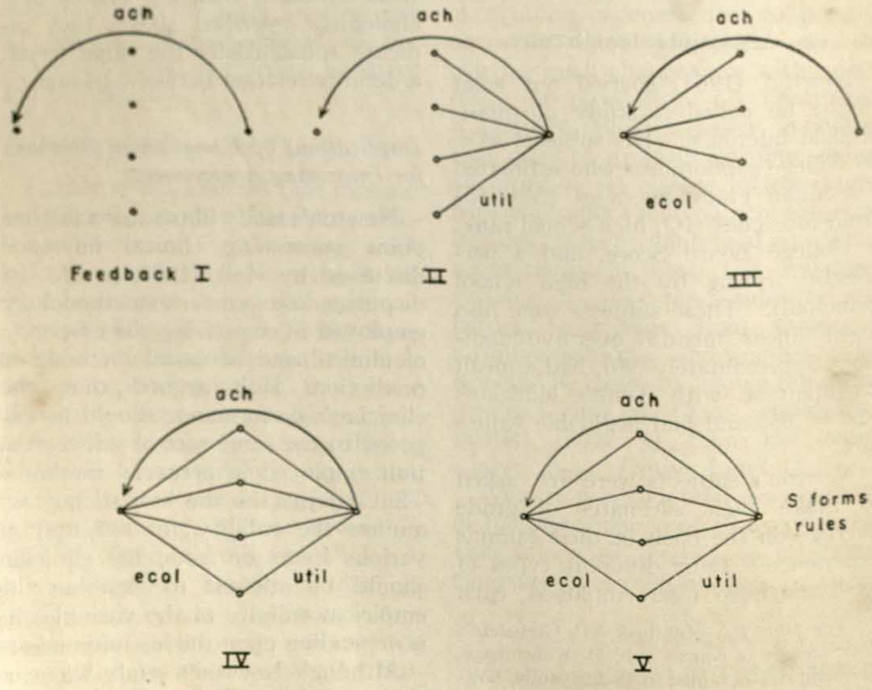


FIG. 2. Different types of feedback provided by Newton (1963).

---

[6] A recently completed experiment by Todd and Hammond (1964) confirms the finding that statistical feedback of this sort is superior to traditional outcome feedback in improving the performance of subjects in a multiple-cue probability learning task.

varying sizes of three geometrical figures on the front of the cards which had probabilistic relations with the number on the back of the cards. The cue validities were .48, .54, and .81. The R_e for the task was .99. Aside from the high R_e, this study is analogous to the task facing the clinicians who must estimate the patient's IQ from Rorschach factors which have probabilistic relations with IQ. It is different, however, in that it is a learning task, and thus similar to the studies by Smedslund (1955), Summers (1962), and Lee and Tucker (1962).

The results from Rappoport's study are taken from 15 randomly selected subjects' performance on the first 50 trials, although the asymptotic level had not been reached for all subjects: (*a*) the median R is .61, again indicating a fairly high degree of linearity on the part of the subjects; (*b*) negligible amounts of C are present, indicating no correlation between nonlinear aspects of the stimulus situation and the subjects' response system—which is as it should be since the R controlling the stimulus arrangement is .99, thereby leaving virtually zero variance remaining in the environment; (*c*) mean $\Sigma d = .649$, somewhat larger than that observed in the above-mentioned studies. (It must be remembered that these subjects were still in an early stage of learning.)

As in the previous studies, we find a substantial correlation between Σd and r_a; at the end of 50 trials it is $-.86$. Because the multiple R controlling the stimulus arrangements was virtually 1.00, it is noteworthy that the rank order correlation between Σd and R_s is $-.70$.

These results make clear the implications of the equation

$$r_a = \frac{R_e^2 + R_s^2 - \Sigma d}{2}$$

for when R_e^2 is virtually 1.00 as in Rappoport's study, then the only way to attain maximum achievement ($r_a = 1.00$) is by maximizing R_s^2 (developing a completely linear response system) and minimizing Σd (matching ecological validities with utilization coefficients). It is for this reason that Rappoport's study is the only one of those discussed here in which there is a high correlation between R_s and r_a. In the other studies, the departure of R_s from unity was sufficiently large to reduce substantially the correlation between R_s and r_a. In these studies, subjects could increase their achievement *either* by raising R_s or reducing Σd; here it was incumbent upon them to do *both*. This illustrates the fact that the nature of the environment determines the appropriateness of various performance characteristics of the response system. In the most fundamental sense, it vindicates Brunswik's (1957) argument that the environment deserves as much consideration as the organism in the effort to understand achievement.

IMPLICATIONS FOR TEST CONSTRUCTION

Cronbach (1956) pronounced "global assessment" to be a failure. He then proceeded to put his finger on the inferential process as the trouble spot: "Assessment encounters trouble because it involves hazardous inferences." And the difficulty is charged specifically to the fact that assessors "attempt a maximum inference from tests [p. 173]." Cronbach (1956) saw the prime hazard to be that,

> Assessors have been foolhardy enough to venture predictions of behavior in unanalyzed situations, using tests whose construct interpretations are dubious and personality theory which has more gaps than solid matter [p. 173].

While we are in general agreement with Cronbach's pronouncement, our analysis of inference leads us to argue that *the very process of test construction itself* must take into account the psychology of inference if the test data are to be used as a basis for inference. This means that experimental psychology will have to be a partner in the enterprise. Use of the lens model, although not mandatory, will expedite matters.

Lens Model, Experimental Psychology, and Clinical Inference

The lens model has an environmental side (to which we applied R_e) and an organismic side (to which we applied R_s). As far as the problem of clinical inference is concerned, each side defines a set of problems and perhaps a division of labor. The environmental side involves the problems of test construction, a problem to which psychometricians and psychodiagnosticians devote their energies with little concern for the cognitive process of the clinicians who must infer the psychological state of the patient from the test data. The organismic side involves the cognitive problems of learning, thinking, and perception, problems to which experimental psychologists devote their energies with little concern for the problem of inferring covert states from a stimulus complex composed of interdependent, uncertain data such as those provided by psychological tests. If the lens model does nothing else, it should make plain the fact that clinical test construction and the experimental psychology of inference provide a research domain to which both can contribute, and from which both can profit.

Those experimental psychologists who are interested in the psychology of inference will be likely to differ with clinical psychologists, however, as to the next step. The former are likely to seek answers by simplifying the inferential tasks studied here; the latter by increasing their complexity. Because our prejudice is with the latter, we will consider some of the steps that can be taken toward increasing complexity. First, however, we must note that requests for increased complexity can be directed toward both sides with equal force. If experimentalists are to be asked to construct and to study problems of greater stimulus complexity, clinicians and test makers can be asked to construct tests in ways which are sustained by more than tradition.

Although the clinical studies reviewed above involve more complex stimulus arrangements than those customarily employed in experimental psychology, they are also clearly vulnerable to the criticism of oversimplification. Estimating the patient's location on a single variable such as IQ is a task which is a priori highly conducive to a linear response system. We are confident that linear response systems will appear in future studies of this type, and that such studies will find small differences between the cognitive processes of the clinician, or any human subject, and the multiple-regression equation. Such similarity, if found to be a fact, would not be unimportant, and a great deal needs to be learned about its implications for learning, thinking, and perception as well as its practical implications for clinical prediction. But it is even more important to discover how human subjects deal with more complex systems. Can the lens model method of analysis be of use with these problems?

The answer is yes, and at least two directions of research are indicated.

The first is concerned with the environmental or test side of the lens model, the second deals with the problem of theoretically oriented inferences.

Cue-Criterion Systems

Little if anything is known about the structure or dynamics of the inferential situations with which clinicians are now confronted in their diagnostic work. Although a clinician may discover from the literature that a given test discriminated one diagnostic group from another ($p < .05$), and may even learn that certain items discriminated between these groups ($p < .05$), this information is of little use to him. For such information does not indicate what the ecological validities are between cues and variable(s) to be estimated, what the reliabilities of the cues are, how much redundant information (cue intercorrelation) is present, how much determinacy there is in the cue-criterion variable system as a whole, or how much and which part of the system is dependably linear and how much and which part is nonlinear. And, of course, the clinician knows little if anything about his own cognitive processes. One result of the failure of clinical research to focus on the structure and dynamics of the clinical task is that research hardly affects clinical practice. The connection between research and function is too remote.

If clinical research were to discover what the cue-criterion structure is for the inference tasks clinicians must cope with, one next step would be to undertake experimental research in the laboratory in order to learn how (and to what extent) human cognitive processes can cope with such multiple-probability cue structures. Such an interchange should lead to research which would rationalize the structure of the clinicians' inferential task. And such an interchange would enable the experimentalist to study more complex cognitive tasks than he has chosen to undertake so far.

Although it is possible with the aid of the computer to construct laboratory analogs of present clinical inference tasks, it would be more to the point to start fresh. Present inferential tasks were not developed with specific consideration given to the cognitive capacities and processes of the humans required to make inferences from the data. Therefore, it can safely be predicted that simulation of current inferential tasks would merely uncover their relative deficiencies. And since enthusiasm for current procedures seems to be on the wane in any event, the discovery of further deficiencies seems beside the point. A fresh start would require that construction of cue-criterion systems and the analysis of cognitive processes be carried out in relation to each other. This suggestion involves nothing more than what would be required of any newly constructed stimulus display, for example, say, a cockpit design for dials and gauges (tests) from which a pilot must make inferences.

But reasons for a fresh start stem not only from the question of the inferential capacity of the clinician in relation to the structure and dynamics of the inferential task, but also from questions concerning the role of theory in clinical (or other) inferences.

Theoretically Oriented Inferences

As Meehl (1960) has pointed out, a purely "dispositional," inductive approach is not the only way to conceive of the clinicians' task.

> An alternative, more exciting (and more congenial to the clinician's self-concept) is to view the clinician's cognitive activity as aiming at the assessment of hypothetical inner states . . . which belong to the domain of theoretical entities [p. 20].

In order to apply the lens model to this situation, which we agree is at once more interesting and more representative of clinical inferences, a double lens model is necessary (see Fig. 3). On the left-hand side of the double lens model a distal trait, or theoretical entity is included which is related to *proximal* traits (Meehl's dispositional terms), which, in turn, are probabilistically related to directly observable cues ($A_1 \ldots A_k$) and ($B_1 \ldots B_k$), as in the single lens model (cf., Meehl, 1960, also Sarbin, Taft, & Bailey, 1960, on probabilism).

On the right-hand side of Figure 3 are the inferences made by the clinician. As in the single lens model, the correlation between t_1 and T_1 indicates the accuracy of the clinician's inference about T_1; the correlation between specific cues $A_1 \ldots A_k$ and t_1 indicates the clinician's dependence on these cues. This dependence may be compared with the ecological validities of the cues (but see above remarks concerning weighting of cues).

Of course, in order for an analogy to be drawn between this model and clinical inference, the clinician must have a theory as to how Y might be inferred from variables T_1 and T_2. Moreover, the theory must be explicit as to how the impalpable T_1 and T_2 are to be inferred from the palpable but probabilistic cue data of each single lens model.

The double lens model provides the means for undertaking the research which is adequate to the complexity of theoretically oriented inferences. As theory encompasses more cues, more proximal traits, and more distal

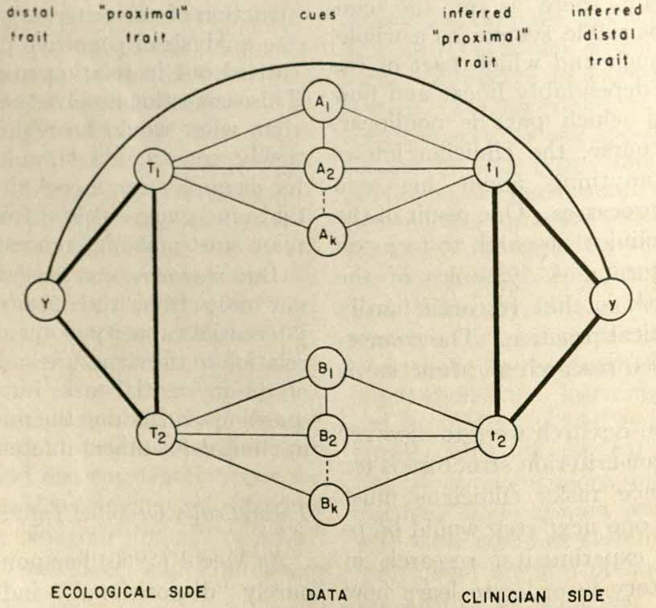


FIG. 3. The double lens model. (Width of line indicates hypothetical degree of correlation.)

traits, then, of course, the double lens model will have to be expanded to a triple lens model, and so on. If this appears to involve more complexity than is desirable, all we need to remember is that no more complexity need be put into the test or environmental side than the organismic side can cope with. At the same time, we need to put *as much* complexity in the environmental side as the organism can cope with if we ever hope to understand the full complexity of the psychology of inference.

The lens model, then, provides a framework for research which places the inferring organism and the inferential situation in their proper relation to one another. Furthermore, it makes possible research which permits a degree of complexity characteristic of the human inference situation while providing a precise analysis of both the situation and the organism's inferential process.

REFERENCES

Beck, S. *Rorschach's test: II. A variety of personality pictures.* New York: Grune & Stratton, 1946.

Brunswik, E. "Ratiomorphic" models of perception and thinking. In N. Mailloux (Ed.), *Proceedings Fourteenth International Congress Psychology.* Montreal, 1954.

Brunswik, E. Representative design and probabilistic theory. *Psychol. Rev.*, 1955, **62**, 236–242.

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

Brunswik, E. Scope and aspects of the cognitive problem. In H. Gruber, R. Jessor, & K. Hammond (Eds.), *Cognition: The Colorado symposium.* Cambridge: Harvard Univer. Press, 1957. Pp. 5–31.

Cronbach, L. Assessment of individual differences. *Annu. Rev. Psychol.*, 1956, **7**, 173–196.

Fink, Carolyn J. Multiple regression analysis of the inferential process. Unpublished master's thesis, University of Colorado, 1961.

Grebstein, L. Relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment task. *J. consult. Psychol.*, 1963, **37**, 127–132.

Guilford, J. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1956.

Hammond, K. Probabilistic functioning and the clinical method. *Psychol. Rev.*, 1955, **62**, 255–262.

Hammond, K. R., Kern, F., Jr., Crow, W. J., Githens, J. H., Groesbeck, B., Gyr, J. W., & Saunders, L. H. *Teaching comprehensive medical care: A psychological study of a change in medical education.* Cambridge: Harvard Univer. Press, 1959.

Hammond, K. Quasi-rationality. Paper read at American Psychological Association, New York, September 1961.

Heider, F. *The psychology of inter-personal relations.* New York: Wiley, 1958.

Hoffman, P. The paramorphic representation of clinical judgment. *Psychol. Bull.*, 1960, **57**, 116–131.

Holt, R. Clinical and statistical prediction. *J. abnorm. soc. Psychol.*, 1958, **56**, 1–12.

Hursch, Carolyn J., Hammond, K., & Hursch, J. Some methodological considerations in multiple-cue probability studies. *Psychol. Rev.*, 1964, **71**, 42–60.

Lee, Joan C., & Tucker, R. An investigation of clinical judgment: A study in method. *J. abnorm. soc. Psychol.*, 1962, **64**, 272–280.

McNemar, Q. *Psychological statistics.* New York: Wiley, 1962.

Meehl, P. *Clinical vs. statistical prediction: A theoretical analysis and review of the evidence.* Minneapolis: Univer. Minnesota Press, 1954.

Meehl, P. The cognitive activity of the clinician. *Amer. Psychologist*, 1960, **15**, 19–27.

Newton, J. Judgment and feedback in a quasi-clinical situation. *J. pers. soc. Psychol.*, 1965, in press.

Rappoport, L. Inter-personal conflict in a probabilistic situation. Unpublished doctoral dissertation, University of Colorado, 1963.

Rommetveit, R. *Selectivity, intuition, and halo effects in social perception.* Oslo: Oslo Univer. Press, 1960.

SARBIN, T., TAFT, R., & BAILEY, D. *Clinical inference and cognitive theory.* New York: Holt-Rinehart, 1960.

SMEDSLUND, J. *Multiple probability learning.* Oslo: Oslo Univer. Press, 1955.

SUMMERS, S. The learning of responses to multiple weighted cues. *J. exp. Psychol.,* 1962, 64, 29–34.

TODD, F. A methodological analysis of clinical judgment. Unpublished doctoral dissertation, University of Colorado, 1954.

TODD, F., & HAMMOND, K. Effect of feedback and set on performance in a multiple-cue probability learning task. University of Colorado, Behavior Research Laboratory Report No. 37. 1964. (Mimeo)

WARD, J. Comments on "the paramorphic representation of clinical judgment." *Psychol. Bull.,* 1962, 59, 74–76.

(Received August 8, 1963)

Psychological Review
1964, Vol. 71, No. 6, 457–472

MOTION AND EMOTION:

THE ROLE OF PROPRIOCEPTION IN THE PHYSIOLOGY AND PATHOLOGY OF THE EMOTIONS[1]

E. GELLHORN[2]

Santa Barbara, California

Proprioceptive discharges contribute to the physiological processes underlying the emotions in two ways: (a) by the setting of the hypothalamic balance which, other conditions being equal, is determined by the posture of the body through the total quantity of proprioceptive impulses impinging on the posterior hypothalamus per unit of time; (b) through facial contraction patterns which lead to afferent discharges via the hypothalamic-cortical system and interact with cutaneous facial impulses in the cortex. It is pointed out that states of abnormal emotional tension are alleviated in various "relaxation" therapies through reducing proprioceptive impulses which impinge on the posterior hypothalamus and maintain the cerebral cortex in an abnormal state of excitation.

> *He who always wears the mask of a friendly man must at last gain a power over friendliness of disposition, without which the expression itself of friendliness is not to be gained—and finally friendliness of disposition gains the ascendancy over him—he is benevolent.*
>
> —FRIEDRICH NIETZSCHE (1876)

There are several lines of evidence for relating the emotional state and the activity of the striated muscles. The psychogalvanic reflex, admittedly a good indicator of emotional tension, is, according to Freeman (1948), "the best general index of the amount of total neuromuscular activity [p. 13]."[3] Total metabolism and activity of the striated muscles are greatly increased during mental work which arouses emotional reactions, but these physiological effects are absent if the same task is carried out without emotional excitement (von Eiff, 1952, 1957). Studies on psychoneurotics show that "progressive relaxation" leads *pari passu* to a decrease in muscle tone and in emotional disturbances (Jacobson, 1938). That anxiety states are associated with increased tension of the skeletal muscles as disclosed by the electromyogram has been pointed out by numerous authors.[4] These data challenge the neurophysiologist to elucidate the underlying mechanisms which seem to be of importance for the therapy of emotional disturbances and for the understanding of the physiology of the emotions. Obviously, the emotions lead not only to autonomic but also to somatic discharges. They induce manifold changes in the tonic and phasic activity of the striated muscles. The former appear as alterations in posture, the latter as contractions of the facial muscles. Both are characteristic for each emotion, and will be discussed from the physiological point of view following an outline of hypothalamic functions.

---

[1] Supported by Grant MH 06552-02 from the National Institutes of Health.

[2] Professor Emeritus of Neurophysiology, University of Minnesota.

Now at 2 Fellowship Circle, Santa Barbara, California.

[3] See also Malmo and Davis (1956).

[4] See Balshan (1962) for the literature.

CHARACTERISTICS OF HYPOTHALAMIC FUNCTIONS

The investigation of autonomic functions by Hess (1949) and others (Gellhorn, 1957b; Ranson, 1940) have clearly shown that the anterior hypothalamus primarily influences parasympathetic activity whereas the posterior hypothalamus regulates sympathetic processes. The physiological significance of these processes becomes obvious in the behavior of the unanesthetized animal in which the autonomic changes are associated with characteristic alterations in the state of the somatic nervous system as indicated by the activity of the striated muscles. Sleep occurring spontaneously or resulting from excitation of the anterior hypothalamus is associated with increased parasympathetic activity (Ban, Masai, Sakai, & Kurotsu, 1951) and relaxation of the muscles whereas in wakefulness or on stimulation of the posterior hypothalamus sympathetic and somatic activity is increased (Hess, 1949; Ranson, 1934). Similarly, on exposure to cold, the sympathetic and the somatic systems (the state of the latter indicated by shivering) are activated whereas on exposure to heat, parasympathetic discharges associated with relaxation of the striated muscles prevail (Ranson, 1940).

In contrast to earlier concepts, according to which the posterior hypothalamus was thought to be responsible solely for the motor expression of the emotions (Masserman, 1943), it is now believed that its excitation elicits processes which underlie also the subjective phenomena of emotion. The reasons are that on hypothalamic stimulation the unanesthetized cat shows a well-directed attack during, and behavioral aftereffects following, stimulation like an emotionally aroused animal (Hess, 1949; Wasman & Flynn, 1962). Moreover, stimulation of the posterior hypothalamus may serve as an unconditioned stimulus (US) since a sensory conditioned stimulus reinforced adequately by the hypothalamic stimulus elicits an emotional response similar to that produced by the US (Affani, Marchiafava, & Zernicki, 1962; Nakao, 1958).

These data suggest that hypothalamic excitation produces not only autonomic-somatic "downward" discharges, but also "upward" discharges to the cerebral cortex. Experiments involving stimulation of the motor area of the cortex showed that hypothalamic stimulation facilitates cortically induced movements (Murphy & Gellhorn, 1945b) and increases pyramidal discharges (Gellhorn, 1957b). Under the influence of stimulation of the sympathetic division of the hypothalamus, "Dial" waves disappear from the cortex and are replaced by more frequent potentials signifying desynchronization while cortical strychnine spikes increase in frequency. These effects are bilateral and reversible (Murphy & Gellhorn, 1945a).

Magoun, Moruzzi, and their collaborators (Magoun, 1954, 1958)[5] discovered that stimulation of the reticular formation in the brain stem (including the hypothalamus, subthalamus, and the reticular thalamic nuclei) produces a similar diffuse excitation of the cortex and, in the unanesthetized animal, arousal, whereas lesions in the medial part of the brain stem induce coma in spite of the integrity of the long ascending afferent system (Lindsley, Bowden, & Magoun, 1949; Lindsley, Schreiner, Knowles, & Magoun, 1950).[6]

---

[5] See also Gellhorn and Loofbourrow (1963), and Rossi and Zanchetti (1957).

[6] The diffuse cortical excitation on stimulation of hypothalamus or reticular forma-

It is believed that tonic afferent impulses from the reticular formation keep the hypothalamus in a reactive state.[7]

The hypothalamus exerts a tonic (excitatory) influence on the cortex since injection of a barbiturate or coagulation of the posterior hypothalamus lessens the state of cortical excitability (Gellhorn, 1957a; Gellhorn & Ballin, 1959; Kennard, 1943; Koella & Gellhorn, 1954; Obrador, 1942). On testing phasic reflexes, it is found that with increasing degrees of excitability of the posterior hypothalamus (induced by intrahypothalamic injection of minute quantities of metrazol or strychnine) reflexes involving the sympathetic nervous system increase in intensity, whereas coagulation of the posterior hypothalamus or local injection of barbiturates exerts the opposite effects (Gellhorn, 1957b; Gellhorn, Nakao, & Redgate, 1956; Gellhorn & Redgate, 1955). Similarly, the degree of hypothalamic-cortical discharges following direct or reflex stimulation parallels the state of excitability of the posterior hypothalamus (Bernhaut, Gellhorn, & Rasmussen, 1953; Gellhorn, 1957a; Gellhorn & Ballin, 1959). These investigations suggest that sympathetic discharges occurring spontaneously or in response to reflex stimulation reflect not only the state of hypothalamic excitability, but also the intensity of the hypothalamic-cortical discharges.[8]

Whereas the ergotropic syndrome consisting of increased sympathetic and motor discharges and diffuse cortical excitation (desynchronization) is elicited by stimulation of the posterior hypothalamus, the trophotropic syndrome is evoked by stimulation of the anterior hypothalamus (Ban et al., 1951; Hess, 1949): Parasympathetic discharges and loss in muscle tone are associated with grouped potentials (synchronization) in the EEG.[9]

Reciprocal relations exist between the two divisions of the hypothalamus as indicated in the behavior of the animal: Anterior hypothalamic lesions lead to sleeplessness and exhaustion (Nauta, 1946) resulting from the release of the posterior hypothalamus, whereas lesions in the latter cause somnolence (Ranson, 1934) suggesting dominance of the anterior hypothalamus.

If through direct stimulation or appropriate reflex stimuli the hypothalamic balance is altered, changes in hypothalamic reactivity occur which are due to reciprocal inhibition. Thus, during dominance of the anterior hypothalamus, the reactivity of the posterior hypothalamus is lessened, and vice versa. These changes in reactivity alter autonomic *and* hypothalamic-cortical discharges (Gellhorn, 1957b).

The reciprocal relation existing between the anterior and posterior hypothalamus (Gellhorn, 1957b; Gellhorn et al., 1956) is demonstrable on heating the anterior hypothalamus in an animal subjected to cold stress. The shivering decreases (Hemingway, Rasmussen, Wikoff, & Rasmussen, 1940) and lessened sympathetic discharges lead to vasodilatation. A similar effect is produced by the electrical stimulation of the anterior hypothalamus (Andersson & Persson, 1957). For

---

tion is abolished by small doses of barbiturates which do not interfere with the conduction of the specific afferent systems to the sensory projection areas (French, Verzeano, & Magoun, 1953; Gellhorn (1953).

[7] See Gellhorn and Loofbourrow (1963) concerning important differences between reticular formation and hypothalamus.

[8] For hypothalamic-limbic relation and further details, see Gellhorn and Loofbourrow (1963).

[9] See Gellhorn and Loofbourrow (1963) concerning other cerebral structures which induce ergotropic and trophotropic effects.

our problem it is of particular interest that these stimuli alter the electrocorticogram and the proprioceptive discharges via the gamma system (von Euler & Söderberg, 1957) at the same time: The cortical potentials become grouped, their frequency is slowed, and their amplitude increased (as in sleep) while the proprioceptive impulses decline in frequency. Apparently the inhibition of peripheral sympathetic and somatic discharges is paralleled by an inhibition of the diffuse discharges from the posterior hypothalamus and the reticular formation to the cortex. Finally, the temperature threshold of the anterior hypothalamus to produce vasodilator action was found to be increased during nociception, which stimulates the posterior hypothalamus, but decreased during stroking of the back (von Euler & Söderberg, 1958), which seems to activate the anterior hypothalamus. The feeling of comfort in warm environment, seemingly displayed also by the cat on stroking the back, is believed to be the result of a shift in the hypothalamic balance to the parasympathetic side, whereas the opposite change in balance resulting in an increased dominance of the ergotropic system on exposure to cold or nociceptive stimuli appears to be related to the feeling of unpleasantness and tension.[10]

AFFERENT IMPULSES AND HYPOTHALAMIC BALANCE

Koch showed in 1932 (Gellhorn & Loofbourrow, 1963) in dogs with exteriorized carotid sinuses that an increased intrasinusal pressure led to a decrease in muscle tone, lessened somatic excitability, and sleep. On the other hand, when the pressure in the carotid sinus was lowered, the general excitability and activity of the animal was greatly increased.[11] Moreover, the knee jerk was found to be inhibited with increasing intrasinusal pressure (Spychala, 1932).

In the light of present knowledge these experiments suggest that variations in intrasinusal pressure alter the ergotropic system and the state of the cerebral cortex through the upward and downward discharges from the posterior hypothalamus and reticular formation. This interpretation is based on the fact that changes produced in the sino-aortic pressure by mechanical means (Bonvallet, Dell, & Hiebel, 1954) or by drugs (Nakao, Ballin, & Gellhorn, 1956) alter cerebral cortical potentials in the expected manner provided that the baroreceptors are intact (Gellhorn, 1959; Nakao et al., 1956). Furthermore, the rise in sino-aortic pressure known to cause increased discharges from the baroreceptors (Heymans & Neil, 1958) causes an increased reactivity of the anterior and a decreased responsiveness of the posterior hypothalamus, whereas the opposite changes accompany a fall in the pressure of the carotid sinus (Gellhorn, 1957b, 1960; Redgate & Gellhorn, 1955).

Sensory stimuli (acoustic < proprioceptive < nociceptive) lead to a diffuse excitation of the cerebral cortex, movements, and sympathetic discharges (Bernhaut et al., 1953). There is a parallelism between hypothalamic and diffuse cortical excitation suggesting that the sensory stimuli induce a hypothalamic-cortical discharge. The failure of nociceptive stimuli to evoke a diffuse cortical excitation in cats with bilateral lesions in the posterior hypothalamus (Koella & Gellhorn, 1954) supports this interpretation. Cold and stress in anticipation of nociceptive stimuli likewise increase the activity of the ergotropic system.

---

[10] Concerning the relation of these phenomena to withdrawal and thigmotactic reflexes, see Ebbecke (1959).

[11] Increase in sino-aortic pressure was shown to decrease experimentally induced convulsions, and vice versa (Gelhorn, Yesinick, Kessler, & Hailman, 1942).

On the other hand, stroking the back elicits parasympathetic effects and synchronized cortical potentials while the muscles relax and the gamma discharges diminish (von Euler & Söderberg, 1958). Stroking also produces muscular relaxation and cessation of crying even in infants without cerebral hemispheres (Edinger & Fischer, 1913) and inhibits cortically induced movements in anesthetized animals (Bubnoff & Heidenhain, 1881). It is suggested that these effects are due to the excitation of the anterior hypothalamus with reciprocal inhibition of the posterior division leading to lessened hypothalamic-cortical discharges.

It seems to follow from the work cited thus far that an increase in the state of activity of the posterior hypothalamus, regardless of whether it is induced by direct or reflex stimuli or whether it results from a shift in the intrahypothalamic balance, leads to a diffuse cortical excitation whereas increased discharges from the anterior hypothalamus are associated with a reduction in the state of cortical excitation. If the striated muscles are not curarized, muscle tone and activity (including proprioceptive discharges) are enhanced on stimulation of the posterior hypothalamus and inhibited in response to the excitation of the anterior hypothalamus. Other conditions being equal, the muscle spindles increase their rate of discharge with increasing activation of the gamma (small nerve fiber) system (Kuffler, Hunt, & Quilliam, 1951). This activation is easily evoked by stimulation of the reticular formation (Granit, 1955) and hypothalamus (Granit & Kaada, 1952). Bearing in mind that proprioceptive impulses are effective stimulants of the posterior hypothalamus, it is concluded that the excitation of the sympathetic division of the hypothalamus in emotion leads to movements reinforced by a positive feedback mechanism.

Since hypothalamic excitation results in a diffuse cortical excitation in completely curarized animals and also in the *encéphale isolé*, the cortical changes are obviously not simply the result of increased proprioceptive discharges. Nevertheless the experiments suggest that in the normal organism such discharges contribute, via the posterior hypothalamus, to cortical excitation. A further proof for this interpretation lies in the following experiments in which the influence of *reduced* proprioceptive discharges on the state of activity of hypothalamus and cortex was studied.

THE INFLUENCE OF BLOCKING PROPRIOCEPTIVE DISCHARGES ON THE STATE OF THE HYPOTHALAMIC SYSTEM

Cats, prepared under ether and local anesthesia (or after intravenous administration of Pentothal) were used several hours later. The posterior hypothalamus was stimulated and the effect of curare-like drugs on this stimulation was recorded on the blood pressure and the nictitating membrane (n.m.). It was found regularly that Intocostrin (administered intramuscularly) in doses which do not alter blood pressure reduces the sympathetic effects of hypothalamic stimulation: The rise in blood pressure and the contraction of the n.m. are lessened. Since the action of Intocostrin is in general too prolonged to demonstrate the reversibility of this action, additional experiments were performed with Anectine (administered intravenously). A similar but reversible reduction in sympathetic hypothalamic responsivness was recorded (Gellhorn, 1958a, 1958b).

In view of the fact that previous work had shown that changes in the excitability of the posterior hypothalamus lead to alterations in sympathetic reactivity as well as to corresponding changes in the intensity of the hypothalamic-cortical discharges, it was expected that these drugs which eliminate proprioceptive impulses would lessen the excitability of the cerebral cortex. It was found, indeed, that the small frequent cortical potentials signifying asynchrony and excitation were replaced by synchronous, grouped potentials (as in sleep) after the administration of Intocostrin or Anectine. The simultaneous recording of the hypothalamic potentials disclosed similar changes. The frequency analysis of these potentials showed that after Intocostrin the amplitude of the integrated potentials was greatly increased for the low frequencies, an effect which is typical for the reduction in excitability and increased synchrony. Finally, it was found that the activation of the hypothalamic-cortical system through reflexes was lessened by Intocostrin. Thus the action of nociceptive stimuli to produce an asynchrony of the potentials of hypothalamus and cortex was reduced in duration and intensity.

These experimental results were confirmed by Hodes (1962) who also produced similar effects with Flaxedil. This drug, which maintains cortical synchrony and a sleeplike state behaviorally (slit pupils!) for hours, failed to alter the EEG when administered intra-arterially. As in the experiments with Intocostrin, it is therefore the lack of proprioceptive impulses inpinging on the central nervous system (due to the blocking of the neuromuscular junction) and not a direction action on the brain which is responsible for the behavioral effects and the action on the cortical potentials of these drugs. As Flaxedil wears off, the cortical potentials become asynchronous, the pupils dilate, and spontaneous movements occur: The reappearance of proprioceptive impulses induces an alert state through the activation of the reticular formation and the hypothalamic-cortical system.

In the normal organism there is a balance between an activating system, which produces cortical asynchrony and increases sympathetic discharges, and a recruiting system, which causes grouped (synchronous) potentials and parasynpathetic discharged to prevail (Tokizane, Kawakami, & Gellhorn, 1957a, 1957b). The weakening of the activating system may be accomplished by reducing or eliminating the proprioceptive impulses. Under these conditions synchrony is produced by subthreshold concentration of barbiturates (Hodes, 1962) or by a subthreshold shock applied to the recruiting system, e.g., via the caudate nucleus.[12] This shift in cerebral balance appears as relaxation, tendency to sleep, and lessened responsiveness to emotional excitation and pain.[13] It seems to account for the success of various relaxation therapies (Jacobson, 1938; Kretschmer, 1949; Schultz & Luthe, 1959) in psychoneuroses and allied conditions.

---

[12] The drug Parpanit, which reduces muscle tone in the spinal animal (Fleisch & Baud, 1948), lowers the recruiting threshold of the caudate nucleus (Tokizane, Kawakami, & Gellhorn, 1957).

[13] With respect to pain, it should be added that the perception of nociceptive stimuli and the reaction to them requires more than the integrity of specific afferent pathways. This is illustrated by the relief from severe pain by frontal lobotomy (Watts & Freeman, 1948), the lessened reaction to pain in animals with lesions in the midbrain reticular formation (Melzack, 1961) and in dogs reared in isolation (Melzack & Scott, 1957). Pain is, apparently, the result of an interaction between sensory impulse and emotional experience. The reduction in the excitability of the posterior hypothalamus lessens emotional responsiveness regardless of whether this effect is due to a loss or reduction of proprioceptive impulses or to the absence of reinforcing stimuli which

POSTURE AND MOOD

Different settings of the hypothalamic balance may be produced by different types and intensities of afferent stimulation. Considering the latter first, it is apparent that the loss in muscle tone which accompanies sadness and similar moods (Klaesi, 1953) is associated with characteristic postures which are diametrically opposed to those seen in a happy mood. The low intensity of proprioceptive impulses prevailing in a state of sadness would seem to contribute to the shift in hypothalamic balance to the parasympathetic side and to its maintenance, and a corresponding statement would apply to the role of increased proprioceptive discharges and the state of sympathetic hypothalamic dominance in some forms of the happy state.

Although a specific stimulus, mostly in the form of symbols (words seen or heard), appears to be the direct cause of mood or emotion, the setting of the hypothalamic balance through the total quantity of proprioceptive impulses impinging on the hypothalamus per unit of time is of considerable importance. If on receiving the news of a great loss, one would make up one's mind to strut back and forth with chest expanded, this posture would interfere with the development of a sad mood appropriate to the occasion. Conversely, sitting with hunched shoulders as if one had just heard of a tragedy would hardly allow one to experience the uplifting effect of listening to a great Mozart quartet. This close relation between bodily posture and mood, recognized in the German language by the word *Haltung*, which means external carriage (posture) as well as the inner attitude, permitted Kretschmer (1949) to make the terse statement: "Die innere Haltung ist von der Aussenhaltung induzierbar und umgekehrt."[14]

Our interpretation of the relation between posture, mood, and emotional responsiveness is supported by experiments involving hypnosis. In numerous tests it has been impossible to induce a certain feeling which is unrelated to the directly suggested motor attitude. If, for instance, the posture accompanying "triumph" is suggested and "locked" in hypnosis, a depressive mood cannot be brought about *unless the postural setting is changed* (Pasquarelli & Bull, 1951).

These experiments and considerations suggest that the mood depends to an important degree on posture, or physiologically speaking, on the quantity of proprioceptive impulses which by their action on the posterior hypothalamus influence the hypothalamic balance.[15] Apparently not only states of muscle tension seen in the neurotic but also physiological variations in the degree of proprioceptive discharges have an important influence on the state of "hypothalamic tuning" and, thereby, on emotional responsiveness.

---

impinge upon the hypothalamus from the cortex in general (Murphy & Gellhorn, 1945) and the frontal lobe in particular (Ward & McCulloch, 1947).

[14] The inner attitude may be induced through the external posture, and vice versa.

[15] At the present state of our knowledge, it seems to be unwise to extend these speculations further except to point out that the happy as well as the tense mood is associated with increased proprioceptive discharges. Conversely, a loss in muscle tone occurs in such widely different moods as those of sadness and postprandial happiness. It is not unlikely, however, that not only the facial expressions but also the patterns of activity of the skeletal muscles are different in these different emotional states.

Needless to say, many other factors such as age and experience not yet amenable to further analysis are responsible for producing the numerous shades of mood known to us and expressed by appropriate verbal symbols (Bollnow, 1956).

Although the hypothalamic balance influences the general type of hypothalamic-cortical discharge which will appear in response to stimulation, it seems highly probable that specific emotions require, in addition, a great variety of stimuli. They may consist of numerous and complex events, but also relatively simple stimulus situations may play a significant role.

FACIAL MOVEMENTS AND EMOTIONS

Since Darwin (1934) and Duchenne (1949), the role of the movements expressing emotions has been studied intensively. The theory of the emotions of James and Lange based on visceral symptoms was rejected by Cannon (1927, 1931) partly on the grounds that these processes, due to their slowness and lack of differentiation could not account for the great variety of human emotions. This argument would hardly hold for the expressive movements of the face. In a recent review Tomkins (1962) emphasizes the following facts: (*a*) the great density and early development of receptor-effector units in the face; (*b*) the dominance of the face in tactual experiments on children; (*c*) the lesser habituation to startle of the muscles of the face than that of other muscles; (*d*) the absence of fascia in facial muscles which makes it possible to contract small bundles of these muscles through the extensive branches of the facial nerves and thereby to increase the variety of contraction patterns.

From the neurophysiological point of view, it seems probable that in the development of an emotion in the social setting the following events play an important role:

1. The hypothalamic discharge, which, initiated by the sight of a smile or the sound of laughter, the threat of an intravenous injection or the sight of blood, etc., leads to alterations in autonomic balance, muscle tone, and cortical excitation.

2. The specific motor patterns of the facial muscles which are similar to those resulting from stimulation of the hypothalamus (Sem-Jacobsen & Torkildsen, 1960) or limbic brain (MacLean, 1955, 1960) and fundamentally different from cortically initiated movements.[16] The former are preserved when, on account of capsular lesions, willed movements of the lower face are abolished (Martin, 1950) and are present also in the newborn in which, due to fetal disturbances, the cerebral hemispheres are lacking (Edinger & Fischer, 1913; Jung & Hassler, 1960).

3. The proprioceptive and cutaneous impulses elicited by tonic and phasic changes in the facial and skeletal muscles.

Sensations as well as stimuli of a higher order involving symbols and memory processes arouse emotions. Murphy and Gellhorn's work (1945 a, b), showing that strychninization of a cortical site located in the sensory projections or association areas leads to the appearance of strychnine spikes in the hypothalamus, illustrates, at least in principle, the underlying neurophysiological mechanisms.[17] Excitation of the hypothalamus causes upward and downward discharges, and proprioceptive discharges resulting from the contraction of any skeletal muscles are enhanced via the gamma system when the brain stem is aroused in emotion.

---

[16] See Hinsey (1940) for the persistence of hypothalamically induced movements after degeneration of corticobulbar and corticospinal fibers.

[17] See also Gellhorn (1957).

Although the activation of the hypothalamic-cortical system through proprioceptive impulses is well established (Bernhaut et al., 1953), no specific experiments have been performed on the relation of the facial muscles to the hypothalamus.[18] In view of their involvement in emotion and their contraction in typical emotional patterns on hypothalamic stimulation, it is to be expected that afferent impulses arising from the face during emotional expression contribute to hypothalamic excitation and hypothalamic-cortical discharges. The great density of the cutaneous receptors in the face and the considerable variety of the patterns of contraction of the facial muscles suggest that the resulting patterns of neocortical excitation and hypothalamic-cortical discharges will match in diversity that of the emotional expressions.

The interaction at the cortical level between impulses originating during emotional expressions in tactile cutaneous end organs of the face and transmitted through specific afferent systems to the sensorimotor cortex and the hypothalamic-cortical discharges originating in proprioceptive discharges from the facial and skeletal muscles have not yet been investigated. It is not unlikely, however, that work on the interaction between hypothalamic and sensory impulses involving optic and acoustic stimuli has furnished results which are applicable to our problem. These studies (Gellhorn, 1954; Gellhorn, Koella, & Ballin, 1954, 1955) have shown that direct stimulation of the hypothalamus or activation of this structure via nociceptive reflexes leads to a marked enhancement of the effect of either optic or acoustic stimuli on the cortex. Thus, direct or reflexly induced hypothalamic excitation which does not seem to alter the electrocorticogram increases the frequency and/or amplitude of the optically induced potentials. Often multiple responses appear in response to the sensory stimuli during hypothalamic excitation, while single potentials prevail in the control test (without hypothalamic involvement). Moreover, the duration of the enhanced response is greatly increased (prolonged after discharges) and even cortical areas which do not respond to the sensory stimulus by itself show marked reactions under the influence of hypothalamic facilitation. The magnitude of these effects is quantitatively related to the intensity or frequency of hypothalamic stimulation, and is particularly well demonstrated when a small area of the sensory cortex is strychninized.[19]

Although the fast conducting muscle afferents (Group I) do not project to the cortex (Mountcastle, Covian, & Harrison, 1952), there is little doubt that movements or muscle tension resulting in excitation of Pacinian corpuscles (McIntyre, 1962), and joint receptors excite the motor cortex (Gay & Gellhorn, 1949; Malis, Pribram, & Kurger, 1953) and modify its reactivity (Gellhorn & Hyde, 1953).

---

[18] It would be of great interest to determine whether distinct patterns of hypothalamic excitation could be aroused through afferent stimulation of various facial muscles, singly and in combinations characteristic for emotional expression.

[19] Further investigations should determine whether cutaneous stimulation of the face tends to excite the underlying muscles as Hagbarth (1952) established for skeletal muscles. If this were the case, the pattern of the proprioceptive facial discharges would probably undergo quantitative and qualitative changes depending on the intensity and spatial distribution of the excitation of the cutaneous area and its relation to the facial muscles. The diversity of the proprioceptive facial patterns impinging on the hypothalamic system in different emotions would thus be greatly increased.

This interaction is thought to underlie perception, and behavioral observations support this interpretation (Gellhorn, 1957). Thus a cat which does not react to a certain optical or acoustic stimulus will follow these stimuli during a mild stimulation of the posterior hypothalamus which by itself results only in pupillary dilation and contraction of the nictitating membranes.[20] But the application of these findings goes further. Experiments involving stimulation of the limbic brain have shown that there is a gradual transition from arousal, during which the animal reacts to environmental stimuli, to an emotional state disclosed by signs of fear and anger (Chapman, Schroeder, Geyer, & Brazier, 1954; Heath, Monroe, & Mickle, 1955; Hunsperger, 1956). It seems not improbable, therefore, that the sensory-hypothalamic interaction in the neocortex accounts also for the qualitative changes in sensations and perceptions during emotional excitation. The fact that this facilitative action of hypothalamic-cortical discharges is not confined to the sensory projection areas and to the motor cortex (Murphy & Gellhorn, 1945) but extends to the association areas as well (Gellhorn, 1954) suggests that it may play a role in recalling past events. It should also be remembered that memory processes as well as perceptions are facilitated in the emotional state (McGinnis, 1949; Miller, 1950; Rapaport, 1942). In the light of this discussion, facial proprioceptive and cutaneous impulses seem to play an important role in facilitating the complex interactions between brain stem and limbic and neocortex which occur during emotion, and to contribute to the variety of cortical patterns of excitation which underlie specific emotions.

THE SIGNIFICANCE OF LOSS OF FACIAL EXPRESSION

The hypothesis which links proprioceptive discharges in general and proprioceptive and cutaneous impulses from the facial area in particular to perception and emotion receives support from unexpected quarters. Clinical psychiatric observations have shown that in paranoid schizophrenia rigidity of facial expression and changes in muscle tone are associated with the complaint of the patient that the persons in his environment behave like marionettes and have masks—as if the loss in emotional expressiveness were associated with a disturbance in perception (von Ditfurth, 1961). This suggestion is corroborated by the finding that in these conditions—in contrast to states of depression—those phenothiazine derivatives are therapeutically effective in which the autonomic (tranquilizing) action is small but the extrapyramidal diencephalic stimulation considerable. The latter leads to symptoms of Parkinsonism in man and to catatonia in the experimental animal (Taeschler & Cerletti, 1959). As von Ditfurth (1961) points out, these drugs seem to modify the relations existing between environment and emotional expression, and determine thereby whether a "normal" or a "paranoid world" results.

Since the emotions from their very beginning—the smile of the infant induced by the mother's smile in the favorable setting of postprandial muscular relaxation—evolve in a certain social milieu, it is perhaps not surprising that the physiological mechanisms underlying the emotions are also of social significance. It remains to be seen whether the fundamental role which von Ditfurth attributes to the expression of the emotions for the maintenance of normal interpersonal relations is justified in man in whom communication is chiefly attained through symbolic speech. Nevertheless, it is of interest to call attention to the fact that caudate stimulation may change social behavior in monkeys although "the motor effect may seem negligible to the observer [Delgado, 1962, p. 221]." Since the extrapyramidal system regulates expressive movements (Jung & Hassler, 1960; Kelly, Beaton, & Magoun, 1946; Magoun, Atlas, Ingersoll, & Ranson, 1937; Martin, 1950), a change in the latter may result from caudate stimulation and may account for the alterations in social behavior. The experimental exploration of this area may bear fruit for physiology, psychology, and psychiatry, and even contribute to an understanding of social relations.

---

[20] See also Thompson (1963) concerning the loss of visual discriminative behavior in rats with posterior hypothalamic lesions.

It has been shown elsewhere (Gellhorn & Loofbourrow, 1963) that the hypothalamic balance plays a central role in the neurophysiological mechanisms underlying the emotional process. Peripheral stimuli as well as excitation of the hypothalamus or of the limbic brain, to mention only a few factors, alter this balance and lead to repercussions in somatically and autonomically innervated peripheral structures as well as in the cerebral cortex. Obviously, the emotions are determined by many factors and only in special conditions may one factor become of critical significance. Although emotions are facilitated through appropriate posture and facial expression, a trained actor knows that he may evoke emotions in the audience without experiencing them. In stressing the modulating rather than the causative role of proprioceptive and cutaneous impulses for the emotions, their practical significance is hardly lessened. The educational and therapeutic value of the control of the expressive movements (including that of the tone of the skeletal muscles) lies in the fact that they may be used to trigger or to inhibit the emotions by the employment of relatively simple physiological procedures.

SUMMARY

In view of the fact that various emotions are accompanied by widely different degrees of tone in the skeletal muscles and characteristic facial patterns, the role of these tonic and phasic changes in the activity of the somatic system for the physiology and pathology of the emotions is discussed. Proprioceptive and tactile impulses are conveyed via the specific afferent systems to the cerebellum and the sensorimotor cortex. In addition, collaterals from the specific afferent systems reach the reticular formation and the hypothalamus and account for the diffuse excitation of the cortex chiefly via the thalamic reticular system.

Physiological experiments show:

1. A characteristic association of somatic and autonomic processes inasmuch as sympathetic discharges and muscle tone undergo parallel changes under physiological conditions, whereas states of increased parasympathetic activity are accompanied by a loss of muscle tone.

2. Increased sympathetic and somatic reactivity results from direct stimulation of the sympathetic division of the hypothalamus or from reflex excitation. The latter is most effective when nociceptive or proprioceptive stimuli are applied or when the pressure in the carotid sinus is diminished.

3. Increased parasympathetic and decreased somatic reactivity result from stimulation of the anterior hypothalamus and also from an increased pressure in the carotid sinus.

4. Experimental evidence suggests that the gamma system of small motor fibers is activated from the posterior and inhibited from the anterior hypothalamus. It contributes to a further increase in proprioceptive impulses impinging on the posterior hypothalamus (positive feedback) when the sympathetic division of the hypothalamus is excited, whereas the opposite processes are evoked on stimulation of the anterior hypothalamus.

5. Excitation of the posterior hypothalamus, directly or reflexly, increases the diffuse excitation of the cerebral cortex, the "upward discharge" paralleling in intensity the sympathetic-somatic "downward discharge." Excitation of the anterior hypothalamus leads to a reduction in cortical excitation (synchronization of cortical potentials).

6. Reciprocal relations exist between anterior and posterior hypothalamus. This statement is valid for the hypothalamic-cortical discharges which determine the state of cortical excitation as well as for the downward discharges.

The application of these principles to the study of proprioception in relation to the hypothalamus shows that increased proprioceptive discharges induced by passive movements increase sympathetic discharges and, via the hypothalamus, the state of alertness (desynchronization of cortical potentials). Moreover, a reduction or inhibition of proprioceptive impulses through blocking of the neuromuscular junction reduces sympathetic and increases parasympathetic discharges, and also diminishes the state of cortical excitation: Synchronous potentials and behavior indicate that sleep has been induced. These experiments are believed to represent a physiological model of the mechanism by which abnormal states of emotional tension are alleviated through various forms of "relaxation" therapy.

It is assumed that proprioceptive discharges elicited by the muscles and the associated excitation of cutaneous sense organs in the skin of the face play an important role in the development of the emotions and a subsidiary role in their reinforcement after they have been established.

Two processes are distinguished:

1. The setting of the hypothalamic balance which, other conditions being equal, is determined by the posture of the body through the total quantity of proprioceptive impulses impinging on the posterior hypothalamus per unit of time.

2. The facial contraction patterns which are characteristic for the expression of particular emotion. These contractions set up complex proprioceptive patterns which arouse the diffuse hypothalamic-cortical system while the sensorimotor area in the cortex is excited via the specific afferent system from tactile and proprioceptive facial receptors. The mode of interaction between the two systems established earlier in Gellhorn's studies on the role of the hypothalamus in perception is believed to involve the sensorimotor and the association areas and to result in a variety of patterns of cortical excitation which qualitatively depend on the configuration of the activated facial muscles and quantitatively on the intensity of the hypothalamic-cortical discharge.

REFERENCES

AFFANI, J., MARCHIAFAVA, P. L., & ZERNICKI, B. Conditioning in the midpontine petrigeminal cat. *Arch. Ital. Biol.*, 1962, **100**, 305–310.

ANDERSSON, B., & PERSSON, N. Pronounced hypothermia elicited by prolonged stimulation of the "heat loss centre," in unanaesthetized goats. *Acta physiol. Scand.*, 1957, **41**, 277–282.

Balshan, I. D. Muscle tension and personality in women. *Arch. gen. Psychiat.*, 1962, **7**, 436–448.

Ban, T., Masai, H., Sakai, A., & Kurotsu, T. Experimental studies on sleep by the electrical stimulation of the hypothalamus of rabbits. *Med. J. Osaka Univer.*, 1951, **2**, 145–161.

Bernhaut, M., Gellhorn, E., & Rasmussen, A. T. Experimental contributions to the problem of consciousness. *J. Neurophysiol.*, 1953, **16**, 21–35.

Bollnow, O. F. *Das Wesen der Stimmungen*, (3rd ed.) Frankfurt: Klostermann, 1956.

Bonvallet, M., Dell, P., & Hiebel, G. Tonus sympathique et activité électrique corticale. *EEG clin. Neurophysiol.*, 1954, **6**, 119–144.

Bubnoff, N., & Heidenhain, R. Ueber Erregungs- und Hemmungsvorgaenge innerhalb der motorischen Hirncentren. *Arch. ges. Physiol.*, 1881, **26**, 137–200.

Cannon, W. B. The James-Lange theory of emotions: A critical examination and an alternative theory. *Amer. J. Psychol.*, 1927, **39**, 106–124.

Cannon, W. B. Again the James-Lange and the thalamic theories of emotion. *Psychol. Rev.*, 1931, **38**, 281–295.

Chapman, W. P., Schroeder, H. R., Geyer, G., & Brazier, M. A. B., Physiological evidence concerning importance of the amygdaloid nuclear region in the integration of circulatory function and emotion in man. *Science*, 1954, **120**, 949–950.

Darwin, C. *The expression of the emotions in man and animals.* London: Watts 1934.

Delgado, J. M. R. Brain centers and control of behavior—animals. In, *The first Hahnemann Symposium on psychosomatic medicine.* Philadelphia: Lea & Febiger, 1962.

Duchenne, G. B. *Physiology of motion.* Philadelphia: Lippincott, 1949.

Ebbecke, U. Physiologie des Bewusstseins in entwicklungsgeschichtlicher Betrachtung. Stuttgart: Thieme, 1959.

Edinger, L., & Fischer, B. Ein Mensch ohne Grosshirn. *Arch. ges. Physiol.*, 1913, **152**, 535–561.

Fleisch, A., & Baud, C. Le point d'impact du parpanit. *Helv. physiol. Acta*, 1948, **6**, 331–345.

Freeman, G. L. *The energetics of human behavior.* Ithaca: Cornell Univer. Press, 1948.

French, J. D., Verzeano, M., & Magoun, H. W. A neural basis of the anesthetic state. *AMA Arch. Neurol. Psychiat.*, 1953, **69**, 519–529.

Gay, J. R., & Gellhorn, E. Cortical projection of proprioception in the cat and monkey. *Proc. Soc. Exp. Biol., N. Y.*, 1949, **70**, 711–718.

Gellhorn, E. The hypothalamic-cortical system in barbiturate anesthesia. *Arch. int. Pharmacodyn. Ther.*, 1953, **93**, 434–442.

Gellhorn, E. Physiological processes related to consciousness and perception. *Brain*, 1954, **77**, 401–415.

Gellhorn, E. Hypothalamic cortical relations, particularly in conditions of hypothalamic imbalance. *Acta physiol. pharmacol. néerl.*, 1957, **6**, 111–128. (a)

Gellhorn, E. *Autonomic imbalance and the hypothalamus.* Minneapolis: Univer. Minnesota Press, 1957. (b)

Gellhorn, E. The physiological basis of neuromuscular relaxation. *Arch. Int. Med.*, 1958, **102**, 392–399. (a)

Gellhorn, E. The influence of curare on hypothalamic excitability and the electroencephalogram. *EEG clin. Neurophysiol.*, 1958, **10**, 697–703. (b)

Gellhorn, E. The influence of the sinoaortic receptors on the tuning of the autonomic nervous system. *Arch. int. Pharmacodyn. Ther.*, 1959, **122**, 221–233.

Gellhorn, E. The alteration of central autonomic excitability and balance induced by noradrenaline and hypotensive drugs (acetylcholine and histamine). *Acta neuroveg.*, 1960, **20**, 490–513.

Gellhorn, E., & Ballin, H. M. Pharmacological studies on hypothalamic-cortical relations. *Arch. int. Pharmacodyn. Ther.*, 1959, **122**, 265–277.

Gellhorn, E., & Hyde, J. Influence of proprioception on map of cortical responses. *J. Physiol.*, 1953, **122**, 371–385.

Gellhorn, E., Koella, W. P., & Ballin, H. M. Interaction on cerebral cortex of acoustic or optic with nociceptive impulses: The problem of consciousness. *J. Neurophysiol.*, 1954, **17**, 14–21.

Gellhorn, E., Koella, W. P., & Ballin, H. M. The influence of hypothalamic stimulation on evoked cortical potentials. *J. Psychol.*, 1955, **39**, 77–78.

Gellhorn, E., & Loofbourrow, G. N. *Emotions and emotional disorders.* New York: Hoeber, 1963.

Gellhorn, E., Nakao, H., & Redgate, E. S. The influence of lesions in the anterior and posterior hypothalamus on tonic and phasic autonomic reactions. *J. Physiol.*, 1956, **131**, 402–423.

GELLHORN, E., & REDGATE, E. S. Hypotensive drugs (acetylcholine, mecholyl, histamine) as indicators of the hypothalamic excitability of the intact organism. *Arch. int. Pharmacodyn. Ther.*, 1955, **102**, 162–178.

GELLHORN, E., YESINICK, L., KESSLER, M., & HAILMAN, H. Carotid sinus reflexes and convulsions. *Amer. J. Physiol.*, 1942, **137**, 396–403.

GRANIT, R. *Receptors and sensory perception*, New Haven: Yale Univer. Press, 1955.

GRANIT, R., & KAADA, B. R. Influence of stimulation of central nervous structures on muscle spindles in cat. *Acta physiol. Scand.*, 1952, **27**, 130–160.

HAGBARTH, K. E. Excitatory and inhibitory skin areas for flexor and extensor motoneurones. *Acta physiol. Scand.*, 1952, **26**, Suppl. 94.

HEATH, R. G., MONROE, R. P., & MICKLE, W. A. Stimulation of the amygdaloid nucleus in a schizophrenic patient. *Amer. J. Psychiat.*, 1955, **111**, 862–863.

HEMINGWAY, A., RASMUSSEN, T., WIKOFF, H., & RASMUSSEN, A. T. Effects of heating hypothalamus of dogs by diathermy. *J. Neurophysiol.*, 1940, **3**, 329–338.

HESS, W. R. *Das Zwishchenhirn*, Basel: Schwabe, 1949.

HEYMANS, C., & NEIL, E. *Reflexogenic areas of the cardiovascular system*, Boston: Little, Brown, 1958.

HINSEY, J. C. The hypothalamus and somatic responses. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1940, **20**, 657–685.

HODES, R. Electrocortical synchronization resulting from reduced cortical drive caused by neuromuscular blocking agents. *EEG clin. Neurophysiol.*, 1962, **14**, 220–232.

HUNSPERGER, R. W. Affektreaktionen auf elektrische Reizung im Hirnstamm der Katze. *Helv. physiol. pharmacol. Acta*, 1956, **14**, 70–92.

JACOBSON, E. *Progressive relaxation*. Chicago: Univer. Chicago Press, 1938.

JUNG, R., & HASSLER, R. The extrapyramidal motor system. In, *Handbook of physiology*. Vol. 2. *Neurophysiology*. Washington, D. C.: American Physiological Society, 1960. Pp. 863–928.

KELLY, A. H., BEATON, L. E., & MAGOUN, H. W. A midbrain mechanism for faciovocal activity. *J. Neurophysiol.*, 1946, **9**, 181–189.

KENNARD, D. W. Effects on EEG of chronic lesions of basal ganglia, thalamus, and hypothalamus of monkeys. *J. Neurophysiol.*, 1943, **6**, 405–415.

KLAESI, J. Über Neurosenlehre und Psychotherapie. In R. Jung (Ed.), *Handbuch der inneren Medizin*. Vol. 5. Berlin: Springer, 1953. Pp. 1260–1319.

KOCH, E. Die Irradiation der pressorreceptorischen Kreislaufreflexe. *Klin. Wchnschr.*, 1932, **11**, 225–227.

KOELLA, W. P., & GELLHORN, E. The influence of diencephalic lesions upon the action of nociceptive impulses and hypercapnia on the electrical activity of the cat's brain. *J. comp. Neurol.*, 1954, **100**, 243–255.

KRETSCHMER, E. *Psychotherapeutische Studien*, Stuttgart: Thieme, 1949.

KUFFLER, S. W., HUNT, C. C., & QUILLIAM, J. P. Function of medullated small-nerve fibers in mammalian ventral roots: Efferent muscle spindle innervation. *J. Neurophysiol.*, 1951, **14**, 29–54.

LINDSLEY, D. B., BOWDEN, J. W., & MAGOUN, H. W. Effect upon the EEG of acute injury to the brain stem activating system. *EEG clin. Neurophysiol.*, 1949, **1**, 475–486.

LINDSLEY, D. B., SCHREINER, L. H., KNOWLES, W. B., & MAGOUN, H. W. Behavioral and EEG changes following chronic brain stem lesions in the cat. *EEG clin. Neurophysiol.*, 1950, **2**, 483–498.

MACLEAN, P. D. The limbic system ("visceral brain") and emotional behavior. *Arch. neurol. Psychiat.*, 1955, **73**, 130–134.

MACLEAN, P. D. Psychosomatics. In *Handbook of physiology*. Vol. 3. *Neurophysiology*. Washington, D. C.: American Physiological Society, 1960. Pp. 1723–1744.

MCGINNIS, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.

MCINTYRE, A. K. Central projection of impulses from receptors activated by muscle stretch. In D. Barker (Ed.), *Muscle receptors*. Hong Kong: Hong Kong Univer. Press, 1962. Pp. 19–29.

MAGOUN, H. W. The ascending reticular system and wakefulness. In J. R. Delafresnaye (Ed.), *Brain mechanisms and consciousness*. Oxford: Oxford Univer. Press, 1954.

MAGOUN, H. W. *The waking brain*. Springfield, Ill.: Thomas, 1958.

MAGOUN, H. W., ATLAS, D., INGERSOLL, E. H., & RANSON, S. W. Associated facial, vocal and respiratory components of emotional expression: An experimental study. *J. Neurol. Psychopath.*, 1937, **17**, 241–255.

Malis, L. I., Pribram, K. H., & Kruger, L. Action potentials in "motor" cortex evoked by peripheral nerve stimulation. *J. Neurophysiol.*, 1953, **16**, 161–169.

Malmo, R. B., & Davis, J. F. Physiological gradients as indicants of "arousal" in mirror tracing. *Canad. J. Psychol.*, 1956, **10**, 231–238.

Martin, J. P. Fits of laughter (sham mirth) in organic cerebral disease. *Brain*, 1950, **73**, 453–464.

Masserman, J. H. *Behavior and neurosis.* Chicago: Univer. Chicago Press, 1943.

Melzack, R. The perception of pain. *Scient. American*, 1961, **204**, 41–49.

Melzack, R., & Scott, T. H. The effects of early experience on the response to pain. *J. comp. physiol. Psychol.*, 1957, **50**, 155–161.

Miller, J. G. The experimental study of unconscious processes. In M. L. Remert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950. Pp. 261–267.

Mountcastle, V. B., Covian, M. R., & Harrison, C. R. The central representation of some forms of deep sensibility. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1952, **30**, 339–370.

Murphy, J. P., & Gellhorn, E. Further investigations on diencephalic-cortical relations and their significance for the problem of emotion. *J. Neurophysiol.*, 1945, **8**, 431–448 (a)

Murphy, J. P., & Gellhorn, E. Influence of hypothalamic stimulation on cortically induced movements and action potentials of the cortex. *J. Neurophysiol.*, 1945, **8**, 341–364. (b)

Nakao, H. Emotional behavior produced by hypothalamic stimulation. *Amer. J. Physiol.*, 1958, **194**, 411–418.

Nakao, H., Ballin, H. M., & Gellhorn, E. The role of the sino-aortic receptors in the action of adrenaline, noradrenaline and acetylcholine on the cerebral cortex. *EEG Clin. Neurophysiol.*, 1956, **8**, 413–420.

Nauta, W. J. H. Hypothalamic regulation of sleep in rats: An experimental study. *J. Neurophysiol.*, 1946, **9**, 295–316.

Nietzxche, F. *Menschliches Allzumenschliches.* Vol. 1. Leipzig: Kröner, 1876.

Obrador, S. Effect of hypothalamic lesions on electrical activity of cerebral cortex. *J. Neurophysiol.*, 1942, **6**, 81–84.

Pasquarelli, B., & Bull, N. Experimental investigations of the body-mind continuum in affective states. *J. nerv. ment. Dis.*, 1951, **113**, 512–521.

Pompeiano, O., & Swett, J. E. Identification of cutaneous and muscular afferent fibers producing EEG synchronization or arousal in normal cats. *Arch. Ital. Biol.*, 1962, **100**, 343–380.

Ranson, S. W. The hypothalamus: Its significance for visceral innervation and emotional expression. Trans. College Physicians, Philadelphia, s. 4, 1934, **2**, 222–242.

Ranson, S. W. Regulation of body temperature. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1940, **20**, 342–399.

Rapaport, D. *Emotions and memory.* Baltimore: Williams & Wilkins, 1942.

Redgate, E. S., & Gellhorn, E. Further investigations on the relation between hypothalamic excitability and the action of hypotensive drugs. *Arch. int. Pharmacodyn. Ther.*, 1955, **102**, 179–186.

Rossi, G. F., & Zanchetti, A. The brain stem reticular formation: Anatomy and physiology. *Arch. Ital. Biol.*, 1957, **95**, 203–405.

Schultz, J. H., & Luthe, W. *Autogenic training.* New York: Grune & Stratton, 1959.

Sem-Jacobsen, C. W., & Torkildsen, A. Depth recording and electrical stimulation in the human brain. In E. R. Ramey & D. S. O'Deherty (Eds.), *Electrical studies on the unanesthetized brain.* New York: Hoeber, 1960.

Spychala, V. Untersuchungen über vegetative Beeinflussung der Muskeleigenreflexe. *Z. ges. exp. Med.*, 1932, **83**, 199–202.

Taeschler, M., & Cerletti, A. Effets de quelques derives de la phenothiazine sur des réactions motrices et émotionnelles du rat. *J. Physiol., Paris*, 1959, **51**, 873–878.

Thompson, J. Development of facial emotion. *Arch. Psychol.*, 1941, No. 264.

Thompson, R. Cortical and subcortical structures mediating visual discrimination habits in the rat. *Bol. Inst. Estud. Méd. Biol.*, Méx., 1963.

Tokizane, T., Kawakami, M., & Gellhorn, E. On the relation between the activating and the recruiting systems. *Arch. int. physiol. Biochem.*, 1957, **65**, 415–432. (a)

Tokizane, T., Kawakami, M., & Gellhorn, E. Pharmacological investigations on the antagonism between the activating and recruiting systems. *Arch. int. Pharmacodyn. Ther.*, 1957, **113**, 217–232. (b)

TOMKINS, S. S. *Affect imagery consciousness.* Vol. 1. New York: Springer, 1962.

VON DITFURTH, H. Zur Problemlage der Pharmakopsychiatrie. In J. S. Achclis & H. von Ditfurth (Eds.), *Befinden und Verhalten.* Stuttgart: Thieme, 1961.

VON EIFF, A. W. Der Einfluss seelischer Belastungen auf Stoffwechsel und Muskeltonus. *Verh. deutsch. Gesellsch. inn. Med.,* 1952, **58**, 468–471.

VON EIFF, A. W. *Grundumsatz und Psyche.* Berlin: Springer, 1957.

VON EULER, C., & SÖDERBERG, U. The influence of hypothalamic thermoceptive structures on the electroencephalogram and gamma motor activity. *EEG clin. Neurophysiol.,* 1957, **9**, 391–408.

VON EULER, C., & SÖDERBERG, U. Co-ordinated changes in temperature threshholds for thermoregulatory reflexes. *Acta physiol. Scand.,* 1958, **42**, 112–129.

WARD, A. A., JR., & MCCULLOCH, W. S. The projection of the frontal lobe on the hypothalamus. *J. Neurophysiol.,* 1947, **10**, 309–314.

WASMAN, M., & FLYNN, J. P. Directed attach elicited from hypothalamus. *Arch. Neurol. Chicago,* 1962, **6**, 220–227.

WATTS, J. W., & FREEMAN, W. Frontal lobotomy in the treatment of unbearable pain. *Res. Publ. Ass. Nerv. Ment. Dis.,* 1948, **27**, 715–722.

(Received August 13, 1963)

Psychological Review
1964, Vol. 71, No. 6, 473–490

DEVELOPMENTAL CHANGES IN PROBLEM-SOLVING STRATEGIES[1]

MORTON W. WEIR
University of Illinois

An analysis of the performance of Ss at a number of age levels, from 3 yr. to 20 yr., in a problem-solving task. Dependent variables include terminal level of correct response, rate of rise to terminal level of response, patterns of response, and performance as a function of a preceding correct or incorrect response. A U-shaped relationship between age and terminal level of correct response is reported, while age and simple patterns of response bear an inverted U relationship to one another. Younger Ss show a more rapid rise to terminal level of response than do older Ss. In addition, differential effects of reinforcement and nonreinforcement are noted among the age groups, and differential changes occur in these effects as the task progresses. Data from a number of other studies are discussed in relation to these findings, and an attempt is made to consolidate various sources of evidence to provide support for the mechanisms which are assumed to underlie certain of the developmental changes reported. Differential growth of the ability to generate hypotheses and employ strategies and the ability to process the information Ss gain from their own responding is suggested to account for the major developmental differences reported.

The psychological literature on problem solving and related topics such as concept formation is vast and appears to be growing exponentially. Attempts to adequately review this literature (e.g., Chown, 1959; Duncan, 1959) have been hampered by the noticeable absence of systematic research programs which attempt to isolate important variables influencing this type of behavior. Even more scarce are developmental analyses of problem solving and related topics. Most studies which include children as a part of their sample use only a limited age range, and the resulting comparisons across ages are sketchy and incomplete (cf. Russell, 1956). There are some notable exceptions, however. Piaget, in a number of works (e.g., Piaget, 1937; Piaget, 1945; Inhelder & Piaget, 1958), has provided modern developmental psychology with invaluable descriptions and speculations as to the nature of intellectual development from early childhood to adulthood. Excellent reviews of his work may be found in Hunt (1962) and Flavell (1963).

The work of the Kendlers (e.g., Kendler & Kendler, 1962; Kendler, 1963) represents a point of view toward developmental research on problem solving considerably different from that of Piaget. The problem situations which Piaget presents to his subjects generally take the form of laboratory versions of real-life situations. The Kendlers, on the other hand, prefer to use less complex situa-

---

[1] This research was supported by a grant from the National Institute of Mental Health, United States Public Health Service, Grant Number M-5577, and by funds from the University of Illinois Graduate Research Board. I thank the directors and parents of Playtime Nursery and Kindergarten and the University of Illinois Child Development Laboratory for their cooperation, and John Thompson and Gerald Gruen for their aid in data tabulation and analysis. I also thank J. C. DeFries for his statistical advice.

tions in an effort to discover basic mechanisms which they presume will also be important determiners of behavior in more complex, real-life situations.

While this paper is not a defense of either of these approaches, the data to be presented here were gathered from a task which fits the Kendler approach much more closely than it does that of Piaget. A considerable number of studies by other investigators have direct relevance to many of the findings which will be reported here. For the most part, however, these studies make use of a variety of tasks and reinforcement conditions, and utilize only a limited age range. As a consequence, any attempt to review them at the beginning of this paper might make the aggregate appear chaotic and unrelated. For this reason, the results and conclusions of the most pertinent of these studies will be discussed later in the context of the results of the present experiment to which they seem most closely related. An attempt will be made to show that there is considerable agreement among the results of these diverse studies, and that they give support, for the most part, to the developmental data and its interpretation which will be presented in this paper.

METHOD

Subjects

The subjects were 290 children, adolescents, and adults ranging in age from 3 years to 20 years. Preschool children were obtained from private nursery schools, and both elementary and secondary school children were obtained from private and public schools. As far as could be ascertained, these children were from neighborhoods consisting of middle- and upper middle-class families. The data to be reported represent, for the most part, the performance of subjects who had been a part of a number of earlier studies, all of which contained groups receiving the same task and instructions (e.g., Gruen & Weir, 1964; Stevenson & Weir, 1959; Stevenson & Weir, 1963). The adult data were collected from college students as a part of an introductory psychology course.

Apparatus

The apparatus has been described in detail previously (Stevenson & Zigler, 1958). It consisted of a yellow panel containing a horizontal row of three knobs. Above the knobs was a signal light and below the knobs was a delivery hole for marbles. The marbles fell into an enclosed plastic container.

Procedure

Subjects were run individually. Each subject was seated in front of the apparatus and was told that he was to play a game, and that when the signal light went on he was to press one of the three knobs. He was also told that if he was correct, a marble would fall into the enclosed plastic container. The subjects were not allowed to handle or keep the marbles, but the younger children were told that if they won enough marbles they would win a prize which they could keep. Since appropriate prizes are difficult to find for adults and adolescents, this part of the instructions was omitted for these groups. It might be argued that the resulting differences in incentive may have affected performance, but so far as the experimenters were able to tell, motivation at all ages was high. All subjects were instructed that the object of the game was for them to win as many marbles as possible. The subjects were given a total of 80 trials each.

For all subjects, only one of the three knobs ever paid off, and that only part of the time. Two reinforcement percentages for the payoff knob were used, 33% and 66%. Choices of the other two knobs never resulted in the delivery of a marble. For each of the two percentage conditions, the subject was reinforced a certain percentage of the times he actually chose the payoff knob. For example, in the 33% condition reinforcement was delivered on 33% of the subject's choices of the payoff knob, rather than on 33% of the total trials in the task. For each subject, one of the three knobs (either left, middle, or right) was designated as the correct knob. Equal numbers of subjects were tested with each of the three knobs. In each percentage condition, subjects from a number of age levels were included. Table 1 shows the mean ages of all the age groups in the two percentage conditions, as well as the number of subjects within each of these age groups.

This is a deceptive task. It appears quite simple on the surface, but in fact has no solution—at least not the kind of solution that most subjects seem to expect (Stevenson & Weir, 1963; Weir, 1962). It will not, for example, allow any solution which will yield 100% reinforcement. It will not even allow accurate prediction of when a marble will be delivered and when it will not, since the partial reinforcement schedule is randomized. Because the task does not provide a solution acceptable to most subjects, it makes possible examination of problem-solving strategies over a long series of trials. The solution which will yield the greatest number of marbles is 100% choice of the payoff knob. This behavior, which will be referred to as maximization, will be seen to occur as a predominant strategy only in certain age groups.

RESULTS AND DISCUSSION

Terminal Level of Response

The per cent choice of the payoff knob during trials 61–80 as a function of age and reinforcement percentage condition is shown in Figure 1. (For the sake of brevity, choices of the payoff knob will be referred to in the remainder of this paper as *correct* responses, regardless of whether or not the choice actually resulted in the delivery of a marble.) Examination of each of the two percentage conditions separately revealed a significant difference among the age groups in both the 33% condition ($F = 6.60$, $df = 6/163$, $p < .001$) and the 66% condition ($F = 2.48$, $df = 5/114$, $p < .05$). Due to unequal and disproportionate subclass numbers, both percentage conditions were not included in a single analysis of variance. A t test comparing number of correct responses made by subjects in the two percentage conditions indicated that the terminal level of response of subjects in the 66% condition was significantly higher than that of subjects in the 33% condition ($t = 8.18$, $df = 288$, $p < .001$).

The terminal level of response across age groups, shown in Figure 1, appears to be U-shaped in both percentage conditions. A curvilinear regression analysis was performed in order to test the linear and quadratic components of the two percentage curves. Since exact age data were not available for all subjects in some age groups, the mean age of each group (estimated from available age data) was used in a method similar to that suggested by Snedecor (1946, pp. 383–384). From this analysis, the variance due to linear and quadratic components was computed, and the resulting F's showed the quadratic component of the 33% curve to be highly significant ($F = 26.28$, $df = 1/167$, $p < .001$). The quadratic component of the 66% curve was also significant ($F = 10.12$, $df = 1/117$, $p < .01$). The linear component of neither the 33% nor the 66% curve approached significance ($F < 1$ in both cases). Also plotted in Figure 1 is the line of least squares derived from the linear and quadratic components of the regression analysis. In both percentage conditions, these least squares lines appear to fit the

TABLE 1

MEAN AGES AND *N*s FOR SUBJECTS IN THE TWO PERCENTAGE CONDITIONS

33%		66%	
$\bar{X}$ age	N	$\bar{X}$ age	N
3.6	10	3.6	10
5.5	27	5.5	35
7.0	31	7.3	20
9.2	15	9.1	15
10.8	26	13.3	10
14.8	26	18.0	30
18.0	35		
Total	170	Total	120

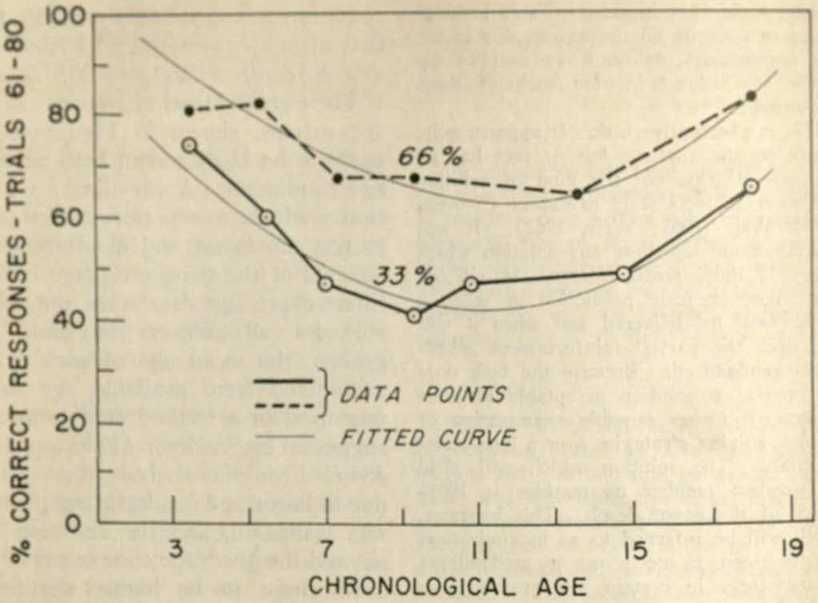


FIG. 1. Per cent correct responses during the last 20 trials as a function of chronological age and percentage of reinforcement. (Both data points and fitted curves are shown.)[2]

data quite well. (For the remainder of the paper, these least squares lines will be referred to as the "fitted" curves.)

One interesting aspect of this U-shaped relationship is that 3-year-olds, 5-year-olds, and college students all show similar levels of performance after 80 trials in this task. There is no question that different processes must have led to this final level for groups so discrepant in age. An examination of the learning curves of the various age groups in the two percentage conditions may provide one indication that markedly different processes are operating.

Group Learning Curves

Figures 2 and 3 present the mean number of correct choices per block of 10 trials for the various age levels in the two percentage conditions. From these two figures it may be seen that the performance of the 3- and 5-year-olds is characterized by a rapid rise to asymptote, while that of the older subjects either shows little change or a slower rise to terminal level of performance. The percentage of subjects in each of the age groups displaying maximizing behavior (i.e., choosing the payoff knob at least 18 out of 20 trials) during the final 20 trials in the task may be seen in Table 2.

One possible explanation for these differences in trends may be the manner in which subjects at different ages

[2] Equations for the two fitted curves are as follows:

33% condition:

$$\hat{Y} = 99.95 - 10.20X + .455X^2$$

66% condition:

$$\hat{Y} = 114.15 - 8.85X + .395X^2$$

where: $\hat{Y}$ = predicted per cent correct

X = chronological age.

approach this task. It is likely that the 3- and 5-year-olds are drawn to the payoff button on the basis of a simple reinforcement notion only. Their tendency to respond to the payoff button is strengthened by each reinforcement, and the tendency to respond to the other two buttons extinguishes rapidly in the face of several nonreinforcements. In order to invoke this explanation of the behavior of younger children, it is necessary to postulate that they do not entertain complex hypotheses involving patterns of response, nor are they particularly disturbed by the partial reinforcement schedule. That is, they may expect neither complex nor "perfect" solutions in a task of this sort, and behave accordingly. A similar result has been reported for children of about this age by Kendler and Kendler (1962). They conclude: "it was possible to infer that as a child matures he makes a transition from responding on the basis of a single unit S-R mechanism to a mediational one." It may be that the single unit S-R mechanisms present in the younger subjects provide an inadequate base for complex strategies. Jones and Liverant (1960) report that younger subjects (4 to 6 years) showed more utilization of "pure" strategy (more maximization) in a two-choice task than did older subjects (9 to 11 years). In addition, their data clearly indicate a more rapid rise in choices of the more frequently reinforced stimulus for the younger than for the

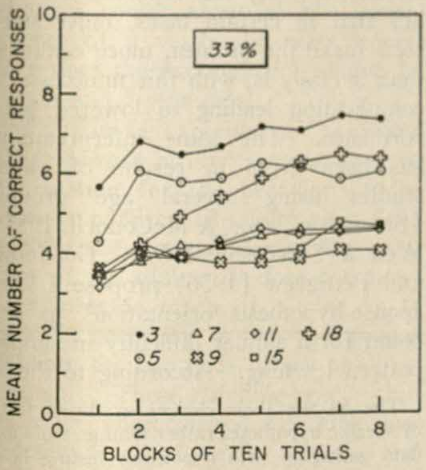


FIG. 2. Mean number of correct responses per block of 10 trials for the seven age groups in the 33% condition.

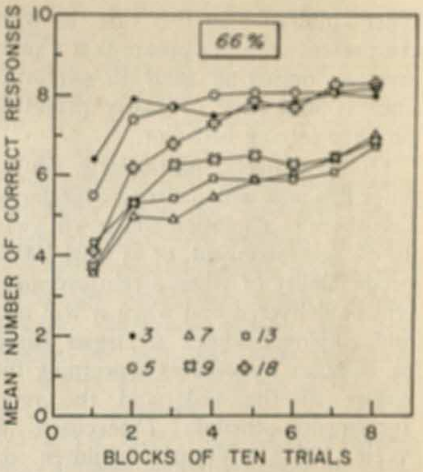


FIG. 3. Mean number of correct responses per block of 10 trials for the six age groups in the 66% condition.

TABLE 2

PERCENTAGE OF SUBJECTS MAXIMIZING DURING THE FINAL 20 TRIALS

33% Condition		66% Condition	
CA	Per cent maximizing	CA	Per cent maximizing
3.6	50	3.6	70
5.5	33	5.5	66
7.0	0	7.3	25
9.2	0	9.1	20
10.8	0	13.3	20
14.8	4	18.0	50
18.0	17		

Note.—The subject was scored as a maximizer if he chose the payoff knob 18 or more times out of the last 20 trials.

older children. In this case, as with the present data, it appears that a pure strategy occurring early in performance is indicative of fairly primitive problem-solving behavior.

Older subjects, particularly adults, enter this task with a strong expectancy that there is a solution which will yield 100% reinforcement, or at least 100% predictability of when a reinforcement will be delivered and when it will not, and employ complex strategies based on complex hypotheses concerning the nature of the task and the reinforcement schedule (Stevenson & Weir, 1963). Younger children, on the other hand, are not concerned with, or more likely not capable of, such complex mediating mechanisms, and do not respond on this basis.

This expectancy for solution that adult subjects show so strongly was encountered by Hyman and Jenkins (1956) who report that it was much more difficult to convince subjects that a sequence was random than it was to convince them that it was structured. A number of investigators have attempted to manipulate this solution set in adult subjects by varying type of task or instructions. Goodnow (1955) found less maximizing behavior in a problem solving than in a gambling (chance) task, and Morse and Runquist (1960) report that matching behavior (nonmaximizing) is more common in a task which the subject believes to have been prescheduled by the experimenter, while more maximizing behavior was noted among subjects who knew that the series was random. A similar finding has been reported by Gruen and Weir (1964) using instructional sets concerning randomness or nonrandomness.

It is interesting to note that in the studies mentioned above, as well as the present experiment, the belief by older subjects that there is a complex solution actually results in fewer choices of the most frequently reinforced alternative. As a consequence, those subjects behaving in the most complex fashion actually receive a lower frequency of reinforcement than do subjects who are behaving in a less complex fashion. The finding that older subjects tend to expect complex solutions, and are hampered in their problem solving by this set, is not without precedent. Kendler (1963) has reported her belief that in certain tasks, older subjects make the problem more complex than it really is, with this unnecessary complication leading to lowered performance. The same interpretation has been given to results of other studies using several age groups (Stevenson, Iscoe, & McConnell, 1955; Weir & Stevenson, 1959). Goodnow and Pettigrew (1956) propose a "response-hypothesis orientation" to account for a similar difficulty in simple pattern learning. According to them,

> The *S*s use their choices as direct tests of specific hypotheses rather than as tools for data gathering with hypothesis testing held in abeyance. As a result, the information gathered is related to a specific hypothesis, and if the latter should prove to be wrong, it is only with difficulty that the information can be transformed and made relevant to another hypothesis. As a rule, *S*s do not transform information but start from scratch again with their next hypothesis [p. 385].

In the present task it appears that even though the adult subjects may have been hampered by such complex hypotheses and strategies early in the task, they were nevertheless able to eventually arrive at a high terminal level of performance. There is little doubt that the frequency of correct responses in this oldest group would have continued to increase had they been given additional trials. It seems reasonable to assume, then, that the older subjects (the 18-year-olds, for example) enter this task with complex hypotheses concerning solution, employ complex strategies, systematically reject these strategies when they do not provide a solution which meets their expectation, and finally arrive at something very close to an actual game solution. That is, they eventually begin to realize that two of the buttons have nothing to do with the game and begin to maximize their choice of the only alternative that ever pays off. But this maximizing strategy comes about as the consequence of a vastly different process than does the maximization shown by the 3- and 5-year-olds.

The data and speculations presented above provide at least some manner of explanation for the behavior of the younger children (3- and 5-year-olds) and for the eventual maximizing tendency of the adults. It does not, however, provide much of a clue as to what is occurring in the age groups between the two ends of the age continuum. These "middle-aged" children are characterized by fairly low terminal levels of correct response and only slight increases in performance across trials (see Figures 1, 2, and 3). More light might be cast on the behavior of this middle age range by examining directly some of the strategies which subjects use in this task.

Response Patterns

Previous experimental findings (Weir, 1962) have indicated that the most common response pattern used in this task (at least for some ages) involves some variant of a left, middle, right (LMR) or a right, middle, left (RML) scheme. This type of pattern represents a simple progressive search across the three response buttons. Figure 4 presents the mean number of such patterns made during the course of the experiment for the various age levels in the two percentage conditions, as well as the fitted curves for the two percentage conditions. Again, due to unequal and unproportional subclass n's, the two percentage conditions were not combined in a single analysis. An analysis of variance performed separately on the two percentage conditions revealed a significant difference among the age groups for both the 33% condition ($F = 4.83$, $df = 6/163$, $p < .001$) and the 66% condition ($F = 5.07$, $df = 5/114$, $p < .001$). A t test comparing mean number of patterns for all 33% groups with all 66% groups indicated that subjects in the 33% condition made significantly more pattern responses of an LMR and RML nature than did subjects in the 66% condition ($t = 6.10$, $df = 288$, $p < .001$).

Again, as with terminal level of performance, a curvilinear relationship between age and performance is indicated. In the case of LMR and RML patterns, however, this relationship takes the form of an inverted U. A curvilinear regression analysis indicated that the relationship between age and performance in the 66% condition is indeed in the form an an inverted U, as the quadratic component was highly significant ($F = 15.45$, $df = 1/117$, $p < .001$), while the linear component did not approach significance ($F = 1.26$). For the 33% condition, however, the linear component was significant ($F = 13.40$, $df = 1/167$, $p < .001$), while the quadratic component did not approach significance ($F < 1$). A Duncan's range test performed on the 33% data revealed the 3-year-old group to differ significantly from the 5-, 7-, and 9-year-olds. This drop at the 3-year level was not sufficient to result in a significant quadratic component, however.

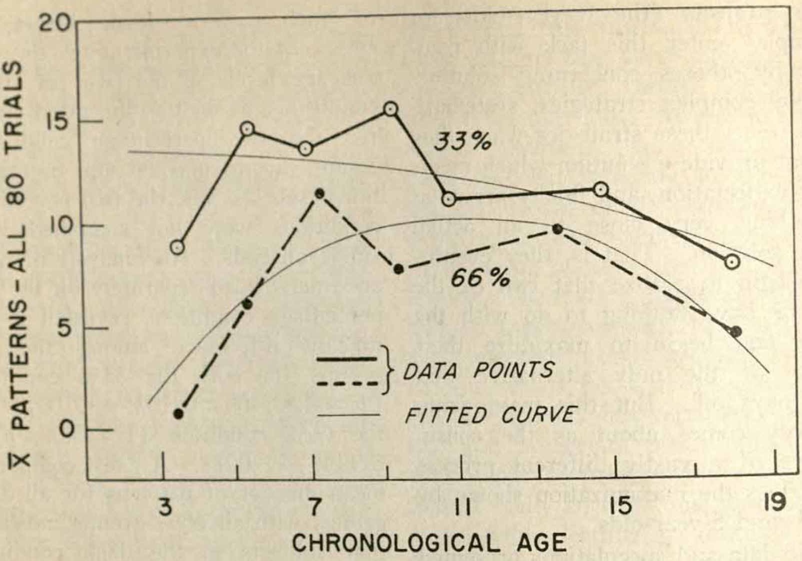


FIG. 4. Mean number of pattern (LMR and RML) responses made during the task as a function of age and percentage of reinforcement. (Both data points and fitted curves are shown.)[3]

From the statistical analysis presented above and from Figure 4, it seems clear that patterns of this simple LMR and RML nature are less likely to occur during the 80 trials in the younger and older groups than they are in the middle age ranges. The number of patterns made as each group progressed in the task was examined, and from this it appears that the 3-year-olds virtually *never* use these types of patterns; the older subjects use them some at the beginning of the task, but they then drop out, and the subjects in the middle age groups tend to use these patterns rather consistently throughout the entire task. This indicates that the very young children do not respond in this patterned way at all, and the older subjects, as was suggested earlier, are able to discard these simple patterns after the first time or two they are unsuccessful.

The LMR and RML pattern data begin to throw some light on the performance of the children in the middle age ranges. Children approximately 7 to 10 years of age respond in a highly stereotyped fashion in this task, employing large numbers of these simple patterns throughout the task. At least two explanations for this stereotyped responding appear possible. Either these middle-aged children do not have more complex patterns to use when the simple ones fail, or they are for some reason unable to reject these simple patterns when they do not pay off and continue to respond in a fairly stereotyped fashion.

---

[3] Equations for the two fitted curves are as follows:

33% condition:

$$\hat{Y} = 13.24 + .144X - .023X^2$$

66% condition:

$$\hat{Y} = -5.97 + 3.00X - .136X^2$$

where:

$\hat{Y}$ = predicted number of pattern responses

X = chronological age.

It is possible that both of these factors are at least partially responsible for the behavior of the children in the middle age ranges. The first explanation, in effect, assumes a severe limitation on the hypothesis formation ability of the child in the middle age range, while the second indicates that this middle-aged child may be capable of complex hypotheses, but he is unable to make full use of the information available from his own responding. This latter explanation would suggest that, in this task, the 7- to 10-year-old is at a point in development where his ability to generate complex hypotheses and employ complex search strategies is growing at a faster pace than his information-processing ability, which catches up only at a later age. Some support for this notion can be found by examining the results of other investigators using similar tasks.

Pattern Responding and Terminal Level of Responses—Data from Other Studies

Most of the information from other studies which can be used for comparison purposes involves decision tasks using only two choices, both of which are partially reinforced. In this type of task, patterns of an LMR or RML nature are clearly impossible, and investigators report that simple alternation is the most frequently encountered response pattern. If alternation in the two-choice task can be looked upon as the analogue of LMR and RML patterns in the three-choice task, then some developmental comparison is possible by combining findings from several recent studies.

Information on alternation behavior is available for the two-choice task for 3- and 4-year-olds from Kessen and Kessen (1961), 5-, 9-, and 13-year-olds from Craig and Myers (1963), 6-, 9-, and 11-year-olds from Gratch (1959), 10-, 13-, and 15-year-olds from Ross and Levy (1958), and some recent data on 3- and 4-year-olds from our laboratory. Although each of these experiments used different tasks, and, in most cases, different reinforcement percentages, it is possible to examine the ordering among age groups within each of these studies as far as alternation behavior is concerned. Combination of these relative orderings may then give some picture of developmental differences in alternation behavior in the two-choice task.

When the data are combined in this fashion, an inverted U function between age and number of alternations is revealed. The shape of this curve is quite similar to that reported in the present study for LMR and RML patterns, with one important exception: the apex of the inverted U for alternations appears to be at about 6 to 9 years of age, while that for the LMR and RML patterns is at 9 to 11 years of age (see fitted 66% curve, Figure 4). If this difference in the apexes of the two curves proves to be a replicable one, it may indicate that as the task gets less complex (reduces from three to two choices), the age at which maximum stereotypy appears also decreases. This would provide support for the notion of differential growth of hypothesis formation ability and information-processing capacity. In other words, the 10- and 12-year-olds are capable, in the two-choice task, of processing the information they are receiving from the outcome of their alternation behavior and consequently reject alternation as a solution. In the three-choice task, however, the increasing complexity involved in series of LMR and RML patterns may take them beyond their capacity to make use of the information they are gaining from their own responses. This inability to make maximum use of their previous response patterns may be mainly a function of insufficient memory for fairly long series of responses and their outcomes which denote the success or failure of such a series. If recall of long series is not possible, subjects may return to the basic pattern in the series for a new attack, and this consistent return may produce the pattern stereotypy noted. If inadequate recall is responsible for this stereotypy, allowing subjects of this age to record and keep track of previous responses and their outcomes should considerably reduce the number of stereotyped response patterns.

If the above analysis is correct, the data from the two-choice task should reveal a U-shaped relationship between age and terminal level of performance, with the lowest point of the curve occurring at a younger age than was the lowest point of the U-shaped curve found in the three-choice task. Combining terminal level of response data from the studies mentioned above, with the addition of a study by Crandall, Solomon, and Kellaway (1961), reveals this to be the case. The lowest point of the terminal level of response curve for the two-choice data appears to be at about 5 to 7 years of age, while Figure 1 shows the lowest point of the two fitted curves representing terminal level of correct response to be at 11 to 12 years of age.[4]

It should be noted at this point that the two response measures which have been discussed thus far—terminal level of correct response and patterns of response—are not necessarily independent of one another. If the subject maximizes his choice of the most frequently reinforced alternative, he will be showing no LMR or RML patterns. Conversely, if he is showing a large amount of pattern behavior, his terminal level of correct response will be low. An attempt was made to take this into account by examining the amount of response variability accounted for by LMR and RML patterns and expressing this as a fraction of the total amount of response variability. In doing this, it was possible to examine only the responses subjects made which were not a part of a maximizing strategy and determine what proportion of those variable responses were taken up by patterns. In other words, if a subject maximized completely during the entire last half of the task, this measure of patterning would take into account only the first half, and would indicate the amount of patterning the subject was showing when he *did* vary his responses.

A variable response was scored any time the subject chose an alternative on Trial $n+1$ which was different from his choice on Trial n. For each subject, the total number of variable responses was computed as well as the number of variable responses which were a part of LMR or

---

[4] Results of a study by Jones and Liverant (1960) appear to be slightly out of line with the findings just presented. They found maximizing by a group of subjects 4.5 years of age, while children of 9 to 11 years of age tended to probability match. The finding of matching in the older children is in accord with the results of Craig and Myers (1963), but the maximizing tendency of the younger children is in contradiction to the findings of Kessen and Kessen (1961), and recent results from our own laboratory. One reason for this discrepancy may be that Jones and Liverant used a task much different from those of the other investigators, which may have affected the performance of the younger children.

RML patterns. When the ratio of variable responses in patterns to total variability was examined, curves nearly identical to those in Figure 4 were the result. This indicates that the pattern of response data presented here does add information to that provided by the terminal level of response data, since the same inverted U relationship between age and performance was found irrespective of the amount of maximization.

Effects of Reinforcement and Nonreinforcement

Thus far, examination of performance in this task has taken into account number of correct responses and patterns of response, and no mention has been made of the effect of a reinforcement or a nonreinforcement on the subjects' behavior. If hypotheses and strategies differ as a function of age, then the effect of a reinforcement or a nonreinforcement should also differ with age. A reinforcement, for example, may be reacted to quite differently by a subject who is in the process of going through a complex pattern than it would by a subject who was not using a pattern strategy. In order to assess the immediate effect of a reinforcement or a nonreinforcement, the performance of subjects on the trial following a reinforcement or nonreinforcement was examined. Each subject was scored, on each choice of the correct knob, as having either repeated that choice on the next trial or having switched to another knob. This tabulation was performed for the payoff knob only.

Several types of strategies may be reflected in these data. For example, a subject who maximized his choice of the correct knob early in the task and continued this maximization would show repeated response to the correct knob following *both* reinforcement and nonreinforcement. This would be termed a win-stay, lose-stay strategy. On the other hand, a subject who was in the process of checking out a fairly complex sequential hypothesis might be virtually unaffected by the occurrence of a single reinforcement or nonreinforcement, and might show any of a number of behaviors, depending on the nature of the sequential hypothesis involved. Other types of strategies such as win-stay, lose-shift, or win-shift, lose-shift, etc., would, if encountered, provide information concerning the effects of reinforcement and nonreinforcement on strategies used in this task.

For each subject, the per cent response repetition was computed for all trials following either a reinforcement or a nonreinforcement. The average per cent response repetition for each of the age groups in the 33% condition is plotted in Figure 5. In this graph, any point above the 50% line reflects a tendency to repeat the previous response, while any point below the 50% line is indicative of a tendency to switch to another choice.

A Lindquist Type I analysis of variance was performed on the data presented in Figure 5. The results of this analysis point up several interesting features of these data. First, the main effect due to age is significant ($F = 7.49$, $df = 6/163$, $p < .001$). There is also a tendency through age 9 for the groups to show more response repetition following a reinforcement than following a nonreinforcement, while at age 11 and beyond, this tendency is reversed. This reversal between ages 9 and 11 is reflected in the analysis of variance by a significant age × reinforcement-nonreinforcement interaction ($F = 6.86$,

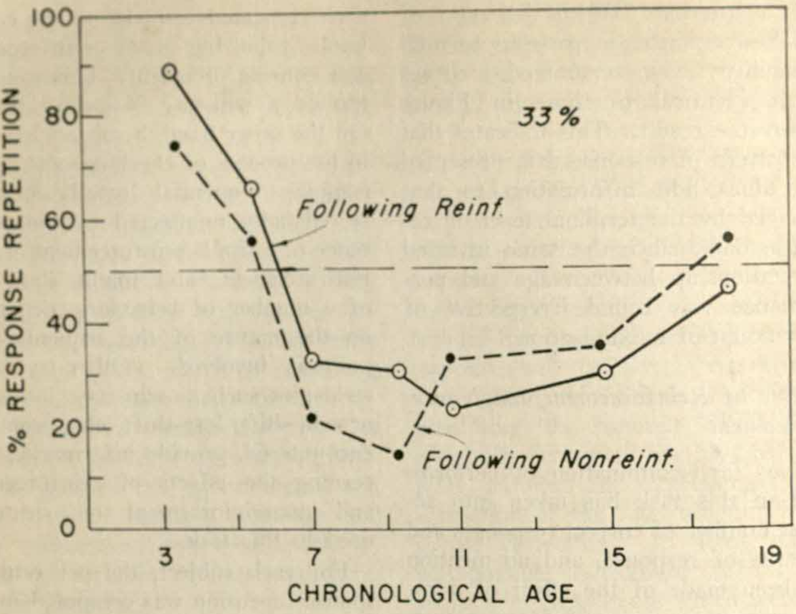


FIG. 5. Per cent response repetition following reinforcement and nonreinforcement for the seven age groups in the 33% condition (all 80 trials).

$df = 6/163$, $p < .001$). The main effect due to reinforcement-nonreinforcement was not significant ($F < 1$).

It appears then, that the younger children (3- and 5-year-olds) show evidence of a maximizing strategy, with a tendency to repeat their previous response regardless of whether it was correct or incorrect. This is especially true of the 3-year-olds. The 7-, 9-, 11-, and 15-year-olds, however, show a tendency toward a win-shift, lose-shift strategy. This probably comes about as a result of their previously noted stereotyped pattern responding, during which the occurrence or nonoccurrence of a reinforcement would have little effect on their behavior. It also indicates that their pattern responding does not contain a large amount of repetition on the payoff knob, but instead is best represented by some sequence of responses involving the other two knobs to a large degree. A strategy employing LMR and RML patterns would fit this definition and yield a win-shift lose-shift outcome.

The greater tendency to repeat the previous response following a reinforcement than following a nonreinforcement for children through age 9 is in agreement with results reported by Brackbill, Kappy, and Starr (1962) and Stevenson and Odom (1963). The reversal in this behavior which occurs at age 11 and older may reflect the fact that older subjects are adopting a rather subtle strategy based on the absence of long runs of reinforcement in the 33% condition. Since this knob pays off on the average of only one out of every three times, following a nonreinforcement a payoff becomes more likely to occur on the next trial. Conversely, if it has just paid off, a nonreinforcement is more likely to occur on the next trial than is a reinforcement. This type of strategy may be too complex to be adopted by the younger children. Goodnow and Pettigrew (1955) report a similar finding with adults trained in a two-choice task with only short runs of reinforcement possible. Their subjects showed a marked tendency to persist with an unsuccessful choice. The same tendency has been reported by Jarvik (1951).

In order to examine changes which might occur in the effects of reinforcement and nonreinforcement as the task progressed, per cent response repetition during the first and last blocks of 10 trials only was computed, and is presented in Figure 6. Although all groups reflect at least some tendency to increase response repetition following both reinforcement and nonreinforcement as the task progressed, the greatest change is shown by the adult group. In order to test for the significance of this and other changes, difference scores were computed for each subject by subtracting the per cent response repetition during the first half of the task from per cent response repetition during the last half. A Lindquist Type I analysis of variance was then performed on these difference scores. This analysis was performed on first and last half of the task rather than on first and last block of 10 trials in order to include a greater number of responses for each subject from which the percentage figures were computed. In the case of the 33% condition, this analysis revealed a significant change in response repetition associated with age ($F = 3.56$, $df = 6/163$, $p < .01$), indicating all ages did not change the same amount during the task. The change in response repetition from first to last half the the task was also different for responses following reinforcement than it was for those following nonreinforcement ($F = 5.12$, $df = 1/163$, $p < .05$). The interaction between age and reinforcement-nonreinforcement was not significant ($F < 1$).

Figure 7 presents percentage of response repetition following both reinforcement and nonreinforcement for the six age groups of the 66% condition. As in the 33% condition, the difference among the age groups was highly significant ($F = 4.29$, $df = 5/114$, $p < .001$). From Figure 7 it can be seen that the tendency to repeat the previous response was greater following reinforcement than nonreinforcement

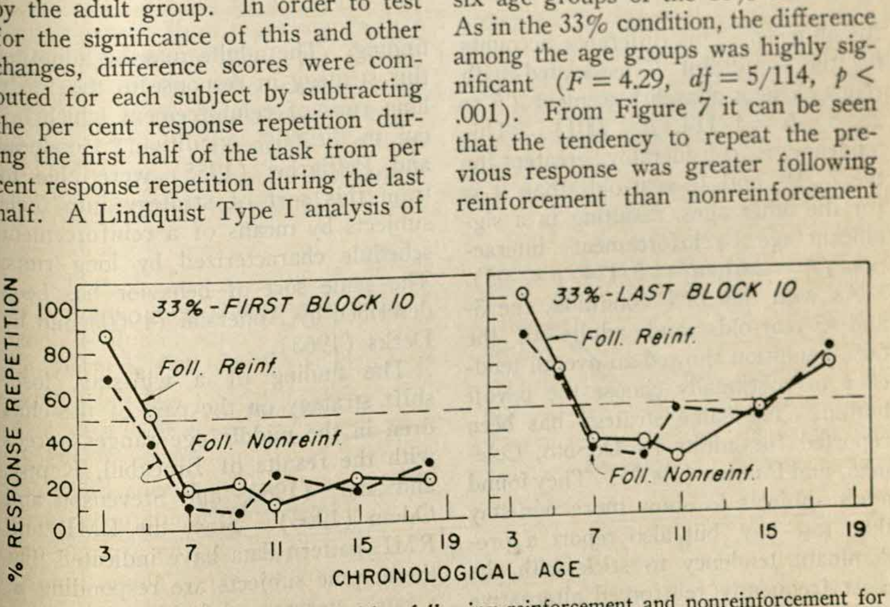


FIG. 6. Per cent of response repetition following reinforcement and nonreinforcement for the seven age groups in the 33% condition. (Graph on left is first block of 10 trials; graph on right is last, eighth, block of 10 trials.)

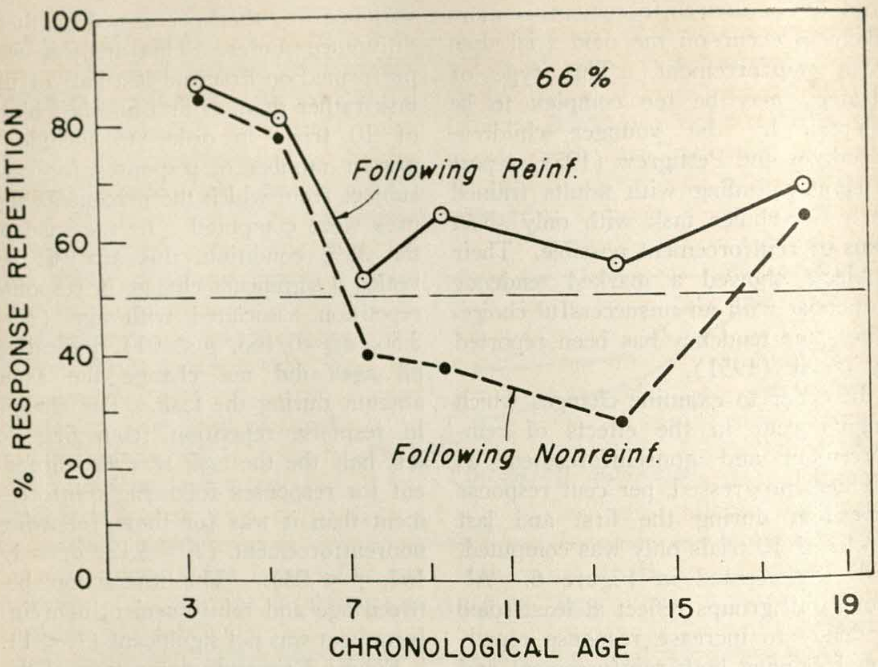


FIG. 7. Per cent of response repetition following reinforcement and nonreinforcement for the six age groups in the 66% condition (all 80 trials).

for all ages. This difference accounts for the significant F associated with reinforcement-nonreinforcement ($F = 28.95$, $df = 1/114$, $p < .001$). This tendency is considerably greater for the 7-, 9-, and 13-year-olds than it is for the other ages, resulting in a significant age × reinforcement interaction ($F = 2.61$, $df = 5/114$, $p < .05$).

As with the 33% condition, the 3- and 5-year-olds and adults in the 66% condition showed an overall tendency to continually choose the payoff button. The same strategy has been reported for adults by DeSoto, Coleman, and Putman (1960). They found adult subjects to show more win-stay than lose-stay, but also report a predominant tendency to stick with the most frequently reinforced alternative after either a win or loss. Hyman and Jenkins (1956) report a similar finding. The adults may be adopting this strategy in response to the fairly long runs of reinforcement which occur in the 66% condition. Goodnow and Pettigrew (1955) were able to train this sort of strategy into their subjects by means of a reinforcement schedule characterized by long runs. The same sort of behavior has been described by Anderson (1960) and by Derks (1963).

The finding of a win-stay, lose-shift strategy on the part of the children in the middle age ranges agrees with the results of Brackbill, Kappy, and Starr (1962) and Stevenson and Odom (1963). Since the LMR and RML pattern data have indicated that these same subjects are responding in a fairly stereotyped fashion, it appears that they not only display pattern stereotypy, but also, during their pattern sequences, tend to repeat the previously reinforced response following a reinforcement, and to shift following a nonreinforcement. A careful check of individual data sheets shows this to be the case. One discrepant finding has been reported by Cohen and Hansel (1955). They report that 6- to 8-year-olds show more win-shift, lose-stay strategy, while older subjects (10 to 11 years of age) were not affected by the outcome of their previous response. The source of this discrepancy in results is not clear.

As with the 33% condition, response repetition following both reinforcement and nonreinforcement in the 66% condition was computed for the first and last block of 10 trials separately in order to examine changes in strategy as the task progressed. These are presented in Figure 8. Again, there is a general tendency for response repetition to increase from first to last block for all ages. Certain of the age groups show more change than others, particularly the 7- and 9-year-olds and adults, who change most drastically in the direction of maximization. This differential change among the age groups is the source of a significant F associated with age in the analysis of change scores from the first to the last half of the task ($F = 2.66$, $df = 5/114$, $p < .05$). No other significanct differences were revealed by this analysis.

One other performance measure seems worthy of comment. In order to assess the degree of tendency toward response repetition which subjects bring with them into this task, the proportion of subjects at each age level who repeated their previous response after the *first* reinforcement or nonreinforcement was calculated. These proportions are plotted in Figure 9. These are first trial data only (or, more precisely, second trial data, since they represent what the subject did following his first reinforcement or nonreinforcement), and represent the sets and tendencies the subjects bring with them to the experimental situation or arrive at through the instructions given. These data present a remarkably similar picture to those presented in Figures 5 and 6, indicating that the 33% reinforcement schedule may not have been sufficient to produce much change in the strategies involving reinforcement and nonreinforcement. It should be noted that the data presented in Figure 9

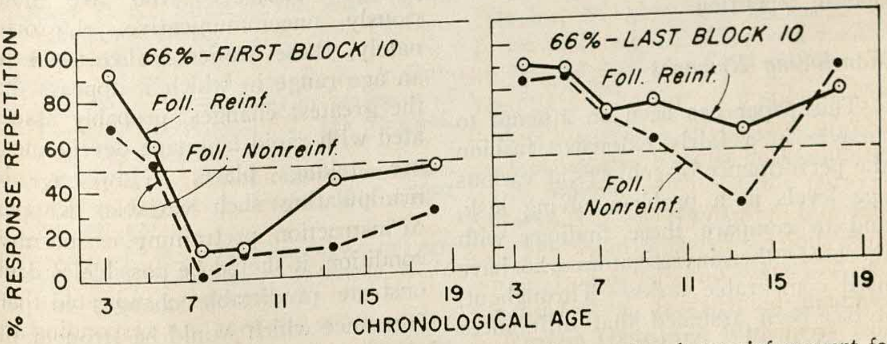


FIG. 8. Per cent of response repetition following reinforcement and nonreinforcement for the six age groups in the 66% condition. (Graph on left is first block of 10 trials; graph on right is last, eighth, block of 10 trials.)

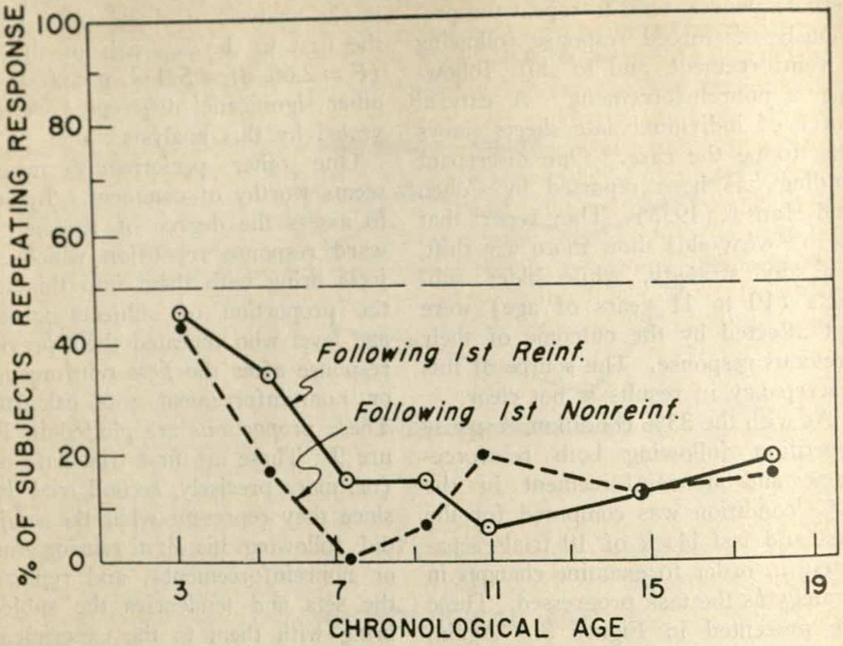


FIG. 9. Percentage of subjects repeating same response following first reinforcement or nonreinforcement. (All percentage conditions combined.)

are not strictly comparable to those in Figures 5 and 6. The data in the latter two figures represent per cent response repetition shown by groups of subjects summed over a series of trials, while the data in Figure 9 are simply proportions of subjects within each of the age groups who show response repetition.

Concluding Remarks

This paper has been an attempt to analyze in a fairly extensive fashion the performance of subjects at various age levels in a problem-solving task, and to compare these findings with those of other investigators who have used comparable tasks. Throughout, it has been assumed that differences in performance at different age levels or at different points during the task reflect hypotheses and strategies on the part of the subjects. Since these hypotheses were not measured directly, they can only be inferred in a general way from the performance data, and their actual existence remains to be demonstrated by further research. This may be impossible for the younger subjects, who are notoriously uncommunicative. Unfortunately, these subjects also represent an age range in which it appears that the greatest changes, probably associated with rapid language development, are taking place. However, by manipulating such variables as type of instruction, pretraining, or incentive condition, it should be possible to demonstrate predictable changes in performance which would be stronger indicators of hypotheses and strategies.

Another limitation to the data which have been presented in this paper is that the strategies discussed were derived from information based on group performance. Obviously, not all subjects within each of the age groups went about solving this problem in the same fashion. However, further research which would include such individual difference variables as intelligence and certain measures of parent-child interaction should make the prediction of individual performance in this task possible.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

BRACKBILL, YVONNE, KAPPY, M. S., & STARR, R. H. Magnitude of reward and probability learning. *J. exp. Psychol.*, 1962, **63**, 32–35.

CHOWN, S. M. Rigidity—A flexible concept. *Psychol. Bull.*, 1959, **56**, 195–223.

COHEN, J., & HANSEL, C. E. M. The idea of independence. *Brit. J. Psychol.*, 1955, **46**, 178–190.

CRAIG, GRACE J., & MYERS, J. L. A developmental study of sequential two-choice decision making. *Child Develpm.*, 1963, **34**, 483–493.

CRANDALL, V. J., SOLOMON, D., & KELLAWAY, R. A comparison of the patterned and nonpatterned probability learning of adolescent and early grade school-age children. *J. genet. Psychol.*, 1961, **99**, 29–39.

DERKS, P. L. Effect of run length on the "gambler's fallacy." *J. exp. Psychol.*, 1963, **65**, 213–214.

DESOTO, C. B., COLEMAN, E. B., & PUTNAM, P. L. Predictions of sequences of successes and failures. *J. exp. Psychol.*, 1960, **59**, 41–46.

DUNCAN, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, **56**, 397–429.

FLAVELL, J. H. *The developmental psychology of Jean Piaget.* Princeton, N. J.: Van Nostrand, 1963.

GOODNOW, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

GOODNOW, J. J., & PETTIGREW, T. F. Effect of prior patterns of experience upon strategies and learning sets. *J. exp. Psychol.*, 1955, **49**, 381–389.

GOODNOW, J. J., & PETTIGREW, T. F. Some sources of difficulty in solving simple problems. *J. exp. Psychol.*, 1956, **51**, 385–392.

GRATCH, G. The development of the expectation of the nonindependence of random events in children. *Child Develpm.*, 1959, **30**, 217–227.

GRUEN, G. E., & WEIR, M. W. The effect of instruction, penalty, and age on probability learning. *Child Develpm.*, 1964, **35**, 265–273.

HUNT, J. McV. *Intelligence and experience.* New York: Ronald Press, 1961.

HYMAN, R., & JENKINS, N. W. Involvement and set as determinants of behavioral stereotypy. *Psychol. Rep.*, 1956, **2**, 131–146.

INHELDER, BÄRBEL, & PIAGET, J. *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures.* (Trans. by Anne Parsons & S. Milgram) New York: Basic Books, 1958.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

JONES, M. H., & LIVERANT, S. Effects of age differences on choice behavior. *Child Develpm.*, 1960, **31**, 673–680.

KENDLER, TRACY S. Development of mediating responses in children. *Monogr. Soc. Res. Child Develpm.*, 1963, **28** (No. 2) 33–48.

KENDLER, H. H., & KENDLER, TRACY S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

KESSEN, W., & KESSEN, MARION L. Behavior of young children in a two-choice guessing problem. *Child Develpm.*, 1961, **32**, 779–788.

MORSE, E. B., & RUNQUIST, W. N. Probability-matching with an unscheduled random sequence. *Amer. J. Psychol.*, 1960, **73**, 603–607.

PIAGET, J. *The construction of reality in the child.* (Orig. publ. 1937) (Trans. by Margaret Cook) New York: Basic Books, 1954.

PIAGET, J. *La formation du symbole chez l'enfant.* (Orig. publ. 1945) [*Play, dreams, and imitation in childhood*] (Trans. by C. Gattegno & F. M. Hodgson) New York: Norton, 1951.

ROSS, B. M., & LEVY, N. Patterned predictions of chance events by children and adults. *Psychol. Rep.*, 1958, **4**, 87–121.

RUSSELL, D. H. *Children's thinking.* Boston: Ginn, 1956.

SNEDECOR, G. W. *Statistical methods.* Ames: Iowa State College Press, 1946.

STEVENSON, H. W., ISCOE, I., & MCCONNELL, C. A developmental study of transposition. *J. exp. Psychol.*, 1955, **49**, 278–280.

STEVENSON, H. W., & ODOM, R. D. Children's behavior in a probabalistic situation. *J. exp. child Psychol.*, in press.

STEVENSON, H. W., & WEIR, M. W. Variables affecting children's performance in a probability learning task. *J. exp. Psychol.*, 1959, **57**, 403–412.

STEVENSON, H. W., & WEIR, M. W. The role of age and verbalization in probability learning. *Amer. J. Psychol.*, 1963, **76**, 299–305.

STEVENSON, H. W., & ZIGLER, E. F. Probability learning in children. *J. exp. Psychol.*, 1958, **56**, 185–192.

WEIR, M. W. Effects of age and instruction on children's probability learning. *Child Develpm.*, 1962, **33**, 729–735.

(Received September 23, 1963)

Psychological Review
1964, Vol. 71, No. 6, 491–504

INFORMATION REDUCTION IN THE ANALYSIS OF SEQUENTIAL TASKS[1]

MICHAEL I. POSNER

University of Wisconsin

This paper proposes a taxonomy of information-processing tasks. Information conserving, reducing, and creating operations are viewed as different methods of processing. The main concern of this paper is information reduction which, it is suggested, represents a kind of thinking in which the solution is in some way implicit in the problem, but in which the input information must be reflected in a reduced or condensed output. A number of tasks within the areas of concept identification and utilization are shown to have this character. If the tasks require complete representation of the stimulus in the response (condensation) the amount of information reduced is directly related to difficulty both during learning and in utilization of previously learned rules. If the tasks allow Ss to ignore information in the stimulus (gating) the direct relation between reduction and difficulty is found during learning but may not occur after the rule is learned.

Most applications of information theory to psychology have been concerned with tasks which require information to be conserved between input and output. In the standard reaction-time task the *S* is supposed to effect an energy and location change, but must preserve all the information in the input for errorless performance (Bricker, 1955a). Verbal-learning experiments add to these requirements the necessity for *S* to serve as a hold circuit and thus delay the response a specified period of time. In studies of the limits of discriminability *S* is required to preserve in so far as possible all stimulus differences in his judgment, the failure to do so is recorded as an error (Miller, 1956). In each of these tasks any change in the informational content during message transmission is clearly error. As useful as these applications have been it is increasingly apparent that this is not the sole nor even the typical information-processing situation for the human. Thus, Bruner (1957) suggests that "the most characteristic thing about mental life is that one constantly goes beyond the information given [p. 41]." With the reverse emphasis, Gerard (1960) suggests that "the real skill of the talented thinker is in discarding irrelevancies. What is omitted in perception, memory and reasoning is of the highest moment [p. 1939]." Attneave (1962) has been somewhat more explicit—"most of the information that goes into the individual never comes out again. The information that is lost is not necessarily wasted, however. The situation is somewhat like that of an executive who considers a mountain of data . . . in order to ar-

---

[1] A portion of this paper is based upon a dissertation submitted to the University of Michigan in partial fulfillment of the requirement of the PhD degree and supported by the Air Force Office of Scientific Research under Contract No. AF 49 (638)-449. The present version was made possible by a grant from the University of Wisconsin Research Committee from funds supplied by the Wisconsin Alumni Research Foundation. The author wishes to thank P. M. Fitts and R. W. Pew for assistance in the presentation of these ideas.

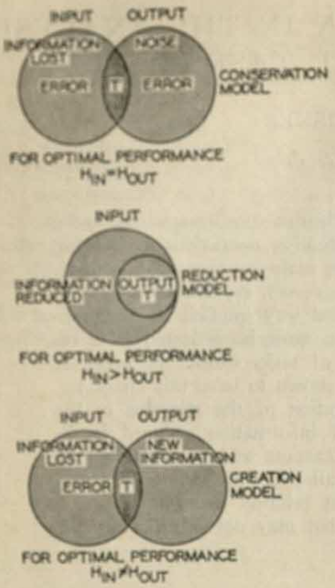


FIG. 1. A proposed taxonomy of information-processing tasks.

rive at a one bit decision [p. 634]." It is with these omissions, condensations, and reductions that this paper will be concerned. It is first useful to outline a general taxonomy of information-processing tasks within which the development of empirical relations can proceed.

TAXONOMY OF TASKS

Figure 1 illustrates a general taxonomy for information-processing tasks which includes not only the commonly studied information-conserving tasks but also information reduction and information creation. This classification considers only the input and output information required for perfect performance. A conservation task such as the standard choice reaction time situation, for example, requires *S* to preserve all the stimulus information for perfect performance. In actual practice he may reduce information in this task by stimulus equivocation, but such a reduction is clearly an error as indicated in Figure 1. In a reduction task, however, *S* is required to map more than one stimulus point into a single response. The loss of information in this situation is not error, but is required by the nature of the task. For example, the sum of a set of numbers has less information than its components, yet in losing this information, *S* has clearly not made an error, but rather has accomplished his task. In the final logical category, the creation task, *S* is required to map a single stimulus point into more than one response. Multiple association is of this character, but the empirical analysis of this type of processing is beyond the scope of this paper. Both the reduction and the creation model seem to capture something of the cognitive aspects of human activity.

Previous Classifications

The necessity for a taxonomy of this sort has become increasingly apparent in the last few years. Studies which place emphasis on many to one (Morin, Forrin, & Archer, 1961) or one to many (Morin & Forrin, 1963) mappings of stimuli into responses have revealed limitations to the generalization so useful in conservation tasks, that task difficulty increases with information transmitted (Bricker, 1955a). Campbell (1958) noted that *S*s tend to show systematic errors in information transmission which may make important additions to the message or which may reduce information through classification. Garner (1962), in his discussion of concept-formation tasks, points out that since these require stimulus equivocation the normal relation between total and internal constraint which exists for conservation tasks is changed. Toda (1963) in reviewing Garner's book recognizes the basic nature of this distinction for task analysis and calls the conservation tasks "normal." Hunt (1962) proposes that learning tasks can be divided into those which require complete information transmission (rote learning), information reduction (concept learning), and information production (probabilistic learning). This type of classification system captures something of Bartlett's (1959) verbal distinction between "closed-system thinking" in which the solution is implicit in the problem and "open-system thinking" in which *S* uses the available evidence to leap beyond the input and provide a creative solution. It has a lesser analogy with the classical division of the thought processes into productive and reproductive (Wertheimer, 1959).

Laws of Information Reduction

A classification system is of little value unless it can be shown that tasks falling within a category obey the same general laws. The study of information-conservation tasks has been useful mainly because performance has been shown to have systematic relations to information transmitted. While these relations are by no means perfect, for example memory tasks show relatively little effect of information per symbol (Miller, 1956), they have provided important basic generalizations from which more specific models have been formulated (Fitts, Peterson, & Wolpe, 1963; Sperling, 1963). This paper proposes that for information reduction tasks difficulty is directly related to the difference between stimulus and response information. The author has chosen to call this difference the information reduced. If stimulus uncertainty is held constant this quantity is the inverse of information transmitted, if response uncertainty is held constant and stimulus uncertainty varied it is, of course, independent of information transmitted. The information reduced is equivalent to amount of stimulus equivocation in conservation tasks, but a distinction is made between equivocation introduced by *S*s' errors and reduction of information required by the nature of the task.

This simple statement of the relation between information reduction and performance allows a variety of experimental situations to be seen as related through the requirement for *S*s to operate on input information to produce a condensed response. Within a restricted task configuration it allows quantitative predictions which will be evaluated in this paper and which should serve to generate new experiments. It is, of course, true that this simple statement does not exhaust our ability to predict, any more than the relation between information transmitted and reaction time completes our knowledge of that task. Within restricted categories more specific models such as the Bourne-Restle model for concept identification may do much better. However, this view provides a basic framework within which more specific formulations can be developed.

Organization of the Paper

This paper will examine first those information-handling tasks which involve the utilization of rules already well learned by *S*. Under this general heading both tasks which allow *S* to ignore aspects of the stimulus (gating) and those that require all aspects of the stimulus to be represented in the response but in a condensed form (condensation) will be considered. The second major area of analysis will concern classification tasks which require the *S* to acquire or learn the use of rules. The paper examines two types of these tasks, first, those which are quite similar to rote learning except that more than one stimulus has the same response (unidimensional) and second, those that meet the full requirements of concept learning by use of multivariate stimulus materials. Within each of these subareas both rules allowing gating and those requiring condensation will be examined.

INFORMATION HANDLING

In attempting to account for compatibility effects in reaction time tasks Fitts (1959) speculated that,

> as the number of intervening information transformations increases, or as the number of hierarchical systems involved in the activity increases, the greater, on the average, is the time required to complete an information-handling task and the greater is the probability of errors [p. 44].

Fitts suggests that such a speculation is not testable by direct observation. This is clearly the case for the information-conserving tasks which he was considering, since the information in the input and output are the same and the experimenter must speculate with respect to intervening transformations. In an information-reduction task, however, the stimulus and response information is different and the experimenter can use the size of transformation between input and output as a directly observable property of the task.

In 1957 Pierce and Karlin studied a task in which *S*s were required either to read aloud a name or to classify it into two categories, male or female in one condition and animal or vegetable in another, by pressing a key. When there were only two names the key press was slightly faster than the verbal response. However, as the population of words increased in size the binary categorization required 40% longer than the naming response.

The information transmitted by the classification task is only 1 bit regardless of the size of the population of individual stimuli, whereas the amount of information transmitted under the information conserving conditions increased monotonically with the number of individual stimuli. The authors conclude that the classification must have taken additional processing time, but did not follow up this inversion of the general rule of increasing reaction time with information transmitted.

Crossman (1953) studied card sorting under a number of conditions. One condition involved sorting the cards into two categories, red and black; a second condition required the *S* to sort into any of four categories, red face, black face, red non-face, black non-face; while the third condition required the *S* to sort red face and black non-face into one category and the reverse into a second category. Each of these tasks involve a loss of information since they require classification into less categories than the information available in the stimulus would allow. The mean sorting time per card for the one *S* studied was .7 second for condition one, 1.1 seconds for condition two, while condition three initially required 1.4 seconds and only after some training approached the speed of condition two. Thus whether transmission of 1 bit is faster or slower than 2 bits depends upon whether *S* is allowed to ignore the face versus non-face distinction of the cards as he is in condition one or whether he is required to process that information in making his classification as he is in condition three.

Gating

We will first look for additional results in reaction time studies which resemble condition one in the Crossman study. Gregg (1954) and Archer (1954) have studied this type of task. In both situations *S*s had to classify one or more binary relevant dimensions, while ignoring one or more binary irrelevant dimensions. In both situations *S*s knew the rules which allowed them to determine whether or not the information was relevant. Gregg (1954) found a small but significant linear increase in reaction time (RT) with the addition of from 1–3 bits of irrelevant information. The size of the increase depended, in part, on the degree of compatibility between the relevant stimulus and its response. Archer, however, found no increase in RT due to the addition of up to two irrelevant dimensions.

In this type of situation *S* can ignore large amounts of stimulus information with little or no penalty in RT. What penalty there is appears to be a linear function of the amount of irrelevant information. Both of these studies require *S* to ignore entire dimensions. When this is the case the way in which the dimensions are separated is probably of importance. For example, studies which present relevant information to the eye and irrelevant to the ear (Broadbent, 1958) and in which *S* knows to attend to the visual input usually show little decrement due to increased irrelevant information. Obviously this would be even more striking with the opposite coding, when *S* could merely shut his eyes as a direct physical gate. The studies of Gregg (1954) and Archer (1954) also involve ignoring entire dimensions, but these dimensions overlap spatially with the relevant ones.

Situations which allow *S*s to reduce information by ignoring aspects of the stimulus will be called gating tasks. The literature indicates that those tasks which allow gating of entire dimensions show little or no increase in difficulty with increasing irrelevant information, once the gating rules are well known. Tasks which require gating within a dimension may show a more marked increase in difficulty with increasing irrelevant information (Rabbitt, 1963) though the evidence here is hardly complete. In a sense the limiting case is composed of those tasks which, like Pierce and Karlin's and Crossman's condition three, do not allow *S* to ignore any aspects of the stimulus, but require him to represent it in a reduced or condensed form. These tasks are conceptually different from the gating tasks, although they still involve information reduction and will be called condensation tasks.

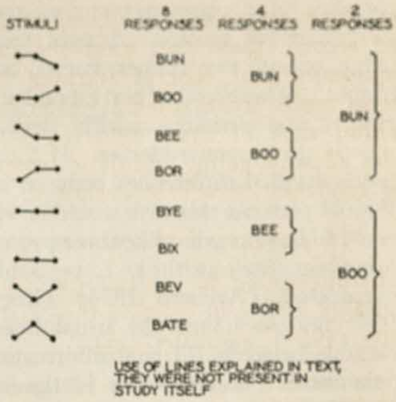


FIG. 2. Stimulus and response sets for reaction-time study (after Bricker, 1955b).

Condensation

Bricker (1955b) presented *S*s with eight stimulus patterns which had to be classified into eight, four, or two response categories. Thus he considered, in addition to the normal information conserving situation, a 1- and 2-bit reduction task. However, the particular coding in the Bricker study (Fig. 2) makes it difficult to be sure whether or not this can be considered a type of gating task. The eight stimuli could easily be recoded, as suggested by the lines inserted in Figure 2, into four perceptually distinct shapes ignoring differences in location and spacing of the elements: curved outward, curved inward, straight, and wedge, or into two shapes, curved and straight. These codes correspond exactly to the divisions which Bricker made in the response classes. If *S* can ignore detailed differences between individual patterns this task would be expected to be somewhat like three conservation situations with 1, 2, or 3 bits transmitted (Archer, 1954; Gregg, 1954) and to follow the usual linear relations between RT and information transmitted. Indeed, this is the result. It is interesting to speculate on what would have happened if all possible codes were used. One might expect that most would provide less opportunity for *S* to ignore stimulus elements in making the classification and the results would become similar to Crossman's condition three and the Pierce and Karlin study. In addition, the Bricker study confounds amount of information reduced with changes in response uncertainty. It would be useful to separate the two effects.

Posner (1962) has used a technique which allows the manipulation of amount of information reduced free from changes in response uncertainty and with systematic variation of the perceptual properties of forms within a response class. Original patterns of eight dots and distortions of these patterns were constructed by means of statistical rules so that a level of uncertainty between an original and its distortions could be stated. The *S* saw two patterns at a time and was required to state whether they were members of the same class or not. Prior to this procedure the *S*s had gone through a rather lengthy learning process where they had learned to give pairs of patterns of varying levels of uncertainty the same names. The results of this learning procedure will be discussed later, but crucial here are the results of the average reaction times when two patterns were exposed which had the same name and when *S* was correct. These data are shown in Figure 3. The results show reaction time to be linearly related to the uncertainty between patterns at least up to 40 bits. The greater the uncertainty which must be smeared over or reduced in making the classification, the greater the classification time.

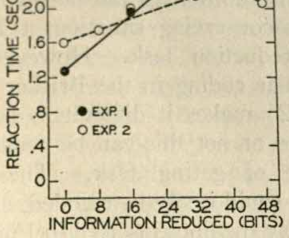


Fig. 3. Reaction time as a function of amount of information reduced.

Morin, Forrin, and Archer (1961) describe an experiment important in this analysis. The stimuli are slides with either one or two squares or one or two circles. Four conditions are crucial: conservation with two stimuli (1 bit), conservation with four stimuli (2 bit), gating in which the four stimuli are classified into circles or squares, and condensation where the response classes are two circles-one square and one circle-two squares. The initial level of difficulty as shown by RT indicates that the gating task was identical with the 1-bit conservation condition, while the condensation task was much slower than these but slightly faster than the 2-bit conservation task. Fitts and Biederman (1964) first replicated the experiment exactly and then added a new stimulus-response (S-R) code which he had found in prior testing to be more compatible. His results closely replicated the original study in the initial difficulty except that with the new S-R code the reduction task took slightly longer than the 2-bit conservation task. In the Fitts results, unlike the original study, the condensation task continues to be of equal or greater difficulty than the 2-bit conservation task over the full 2 days of training.

The results reviewed so far indicate that a condensation task will be much more difficult in terms of processing time than the conservation task with the same information transmitted (Crossman, 1953; Morin, Forrin, & Archer, 1961; Pierce & Karlin, 1957), and will, after training, approach about the same level of difficulty as a conservation task with the same stimulus uncertainty (Crossman, 1953; Fitts & Biederman, 1964). Gating tasks, however (Archer, 1954; Bricker, 1955b; Crossman, 1953; Morin, Forrin, & Archer, 1961) show little or no increased difficulty over their conservation counterpart with the same information transmitted. Only in the Posner (1962) study, in which stimulus similarity is defined in terms of uncertainty, has a quantitative relationship been discussed between the amount of information reduced and reaction time for a condensation task. In that study information was varied by systematic increases in stimulus uncertainty with response uncertainty constant.

Pollack (1963) has developed the basic paradigm of Pierce and Karlin (1957) into a systematic study of the relationship of amount of information reduction to speed of classification. Pollack used 24 generic word categories each having up to 24 individual instances, such as goat and pig as instances of animal. In the main conditions of the study *S*s were required to classify 48 instances into from 2 to 24 categories. When there is only a single instance per category the task is information conserving. The results of these conditions show the expected linear relation between information transmitted and classification time (Bricker, 1955a). However, with the number of response categories fixed there is also a direct and reasonably linear relation between amount of information reduction, from 1 to 4.5 bits per category, and classification time. The slope of this function increases as the number of response categories increases. At low levels of response uncertainty the effect of going from one instance per category (conservation) to two instances (condensation) is much greater than the general straight line trend due to amount of reduction, but the reverse is true at higher levels of uncertainty.

Posner (1962) has performed a study which also allows the production of quantitative relationships between amount of information reduction and performance where stimulus uncertainty is constant and response uncertainty varied. In these experiments the stimulus message was always the same, consisting of a series of eight numbers selected randomly from a population of 1 to 64 and presented aurally to *S*. The *S* was required either to record the numbers or to operate upon them by a number of information-reducing operations. The tasks were chosen so that the output information varied from 7.7 to 48 bits and so that the components of the operations involved were familiar. The tasks included an information-conserving recording condition, a task of alternating recording with summing the digits of a number, a partial addition task in which successive pairs of numbers were added together, a 2-bit classification task where the numbers were classified into high and odd, low and odd, high and even, or low and even, and a 1-bit classification task in which

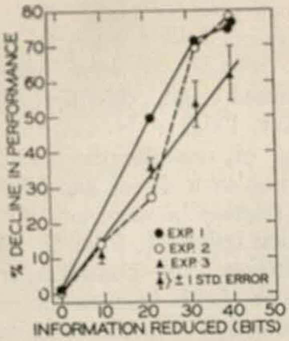


FIG. 4. Percentage decline in performance with speeding as a function of the amount of information reduced.

high and odd or low and even was an "A" and low and odd and high and even was a "B." It was not possible to compare the tasks directly because errors were so different for each task. Therefore, the rate of change of errors in each task was measured as the interstimulus interval was reduced from 4 to 1 second. The basic notion was that the more difficult the task the greater would be the rate of decline of performance with speeding. The use of an error measure, well-learned responses, and an appropriate correction for chance levels were designed to reduce the effect of changes in response uncertainty per se.

Three separate experiments involved the following designs: (*a*) each *S* takes all tasks at one speed, (*b*) each *S* takes one task at one speed, (*c*) each *S* takes one task at two speeds. The results showed that in every experiment the rate of decline in performance increased monotonically with increasing information reduction. The overall relationship, as shown in Figure 4, is roughly linear with some variation between experimental designs. In general, the designs which successfully controlled for sequential transfer effects between the tasks produced more linear relationships.

These results again demonstrate that condensation tasks, unlike conservation tasks, cannot usefully be thought of as increasing in difficulty with information transmitted. On the contrary, when stimulus information is held constant and *S* is required to operate upon it to produce a condensed output as in this last study, difficulty declines with increasing information transmitted. If stimulus information is increased with response information constant as in the tasks measuring reaction time to patterns (Posner, 1962) or in the classification of words (Pollack, 1963) then information transmitted is fixed, but task difficulty increases. In all these cases, however, task difficulty is found to be an increasing linear function of amount of information reduced.

One problem with many of the studies cited here is that they do not adequately control the degree of learning. In all cases the operations or classifications were well known by *S*s but most studies cannot show that they were equally well known or that learning effects were unimportant during the experimental conditions. It is, therefore, important to look at the effect of information-reducing tasks in a situation where the process of learning rules is observed. This will be discussed in the next section.

CLASSIFICATION LEARNING

Unidimensional Rote Learning

Rote learning, like the standard reaction time task, requires information conservation. The minimum requirement for changing a rote-learning task into one of classification learning is to have more than one stimulus per response. When the stimuli lie along a single dimension the *E* has minimal control over the relevant and irrelevant aspects of the stimulus, but he can control the number of stimuli to be assigned a single response and the way in which the stimuli within a response category are grouped along the dimension.

Number of Stimuli per Response. One way to vary the amount of information reduction in a classification task is to increase the number of stimuli per response. Most investigators who have attempted this have kept the number of stimuli constant and reduced the number of responses. Hake and Erikson (1955) and Bricker (1955b) studied the learning of patterns of lights when eight, four, and two response categories were used. Richardson (1958) studied the assignment of 16 adjectives to two, four, or eight response categories. In none of these studies were significant differences found in the rate of learning as a function of the number of stimuli per response, once correction had been made for differing chance levels.

There are several problems in studying the information reduction involved in unidimensional rote concepts. First, all of the studies cited confound the amount of information reduction with the number of response terms to be learned and used. Despite this, the results do indicate that an information-reduction task is more difficult to learn than the conservation task with the same information transmitted, but about the same difficulty as a conservation task with the same stimulus uncertainty. Thus, when *E*s look at the classification of a fixed list of stimuli into a variable number of responses little difference is found among conditions. This is in agreement with the results in many reaction-time studies (Crossman, 1953; Fitts & Biederman, 1964; Morin, Forrin, & Archer, 1961). Future studies should run control groups with conservation tasks of the same response uncertainty as the reduction tasks as well as at the same stimulus uncertainty. Second, these studies allow no clear distinction between gating and condensation. As pointed out earlier, Bricker (1955b) used classifications which allow the opportunity for recording rules which could eliminate processing of certain aspects of the stimuli. It is well known that verbal synonyms (Richardson, 1958) are also subject to such mediational rules. However, in these experiments it is not possible to tell if *S*s are using such rules or when in the learning they come into play. Finally, the most serious difficulty with these studies is that they do not take into consideration the relationship between the stimuli to be assigned a single response. That is, both the number and the way in which stimuli are grouped should effect the total information reduction involved.

Stimulus Grouping. That the number of stimuli per response can be a potent factor when there is more control over coding within a stimulus grouping is shown in a recent study (Smith, Jones, & Thomas, 1963). This study used eight stimuli varying only along the hue dimension. There were either two, four, or eight response categories. When stimuli were assigned at random to response categories the opportunity for recoding by use of familiar color names was minimal. In this situation performance declined uniformly with increasing information reduction. This is in agreement with the general relation between information reduction and performance, even though as information reduction increased the number of responses to be learned declined. If adjacent colors were assigned the same response the task becomes one of absolute discrimination and performance increases as fewer categories are required (Miller, 1956). This study indicates that both the total information reduction and the way information is distributed within response categories is important to predictions of task difficulty. One can vary the number of stimuli per response, but if the added stimuli are very similar or if they can be coded by a well-learned name, little or no change is made in task difficulty.

Some investigators have varied the informational properties of the task by manipulating the variability of the stimuli within a response category. Richardson (1958), using meaningful adjectives, found that as the similarity of items within a response category increased learning greatly improved; the reverse occurred with increases in similarity between categories. Shepard and Chang (1963), using colors previously graded in similarity by use of paired comparisons, assigned subsets of four stimuli to the same response category. They found that the more similar the subset assigned to a response category the faster the learning and that such differences in within response variability accounted for 78% of the learning variance.

French (1953) systematically studied the effect of stimulus similarity within and between response categories. He varied the similarity between the six visual forms assigned to each of two categories by varying the number of line segments which were identical. In the discrimination situation only half of the set of six similar figures were in the same response category, while in the generalization situation they were all in the same response category. These results show that performance is directly related to the degree of similarity within a response category and inversely related to similarity across different categories.

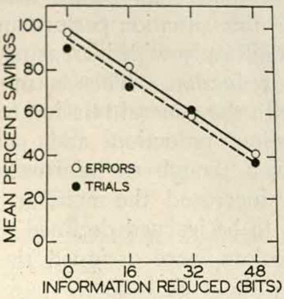


FIG. 5. Rate of transfer as a function of amount of information reduction.

Posner (1964) has shown that it is possible to express the degree of similarity between patterns in terms of uncertainty. Psychophysical studies showed a linear relation between the amount of uncertainty and perceived distance for patterns of eight dots. Such patterns were used in a paired-associate study in which *S*s had to transfer the name learned for one pattern to a new pattern at a given level of uncertainty. It was found that the degree of uncertainty between patterns was linearly related to rate of transfer (Fig. 5).

The result of this study is crucial for three reasons. First, it shows that rate of learning is a decreasing linear function of amount of information within a response category. This is in agreement with studies of stimulus similarity cited above, as one would predict from the linear relation between uncertainty and perceived similarity, and with the results of reaction time studies cited earlier. Second, it allows future studies to combine the number of stimuli per response and uncertainty between stimuli to obtain overall relation between within category uncertainty and rate of learning. Third, it shows how studies of unidimensional classification learning can be systematically compared with multivariate concept-learning tasks which have been explored in informational terms. These relationships will be discussed further in the next section.

Multivariate Concept Learning

The studies which we have been discussing form a bridge between the investigation of rote learning and the investigation of traditional concept formation. They are similar to the concept-formation design in that a single response is made to more than one stimulus. However, they lack one attribute of most concept formation studies (Garner, 1962) namely, they are not multivariate in nature. Garner says,

> this multivariate requirement . . . makes it possible for the experimenter to determine a priori which variables are relevant to changes in the response and which variables are irrelevant to these changes. Then the *S* is given the task of determining which variables are relevant and once he has learned this fact he can respond correctly to a large group of stimuli with a single response by ignoring stimulus differences which exist with respect to the irrelevant stimulus variables [pp. 311–312].

In these types of studies it becomes possible to distinguish rather clearly between gating and condensation.

Metzger (1958) studied paired-associate learning of patterns which had either three or four distinct elements. Each element was a triangle which could be either large or small. In the three element condition he studied conservation tasks with eight and four stimuli, a gating task with four response categories (one irrelevant element) and a condensation task with random assignment of the eight stimuli to four responses. The results showed that the gating and condensation tasks were not much different from the conservation task with the same stimulus uncertainty though much more difficult than the conservation task with the same response uncertainty. This is in agreement with most of the unidimensional results and agrees with the reaction time results except with respect to the gating task. In the four element conditions it is possible to make some quantitative comparisons of the effects of increasing information reduction on rate of learning for both gating and condensation operations. This is shown in Table 1. The conservation condition is for eight stimuli and eight responses, the 1-bit reduction point represents tasks with eight stimuli and four responses, while the 2-bit reduction point represents 16 stimuli and four responses. The results show that both types of reduction tasks are somewhat more difficult than the conservation task with identical stimulus uncertainty and that both increase in difficulty with increasing information reduction.

Gating. Information measurement has been systematically applied to the investigation of concept identification with multivariate stimuli (Archer, Bourne, & Brown, 1955; Bourne & Haygood, 1959; Walker & Bourne, 1961). These studies are gating tasks since the *S*s have to learn to ignore certain irrelevant aspects of the stimulus and then classify all levels of relevant information into distinct categories. The results of these studies show that errors to criterion are an increasing linear function of the amount of irrelevant information which must be ignored in making the response

TABLE 1

MEAN TRIALS TO CRITERION (AFTER METZGER, 1958)

	Information reduced (bits)		
	0	1	2
Gating	21.2	25	38.4
Condensation		40.4	51.2

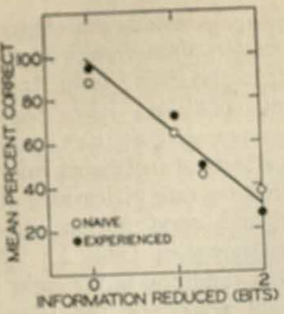


FIG. 6. Percentage of correct responses in concept formation as a function of information reduction required (after Shepard, Hovland, & Jenkins, 1962).

classification. This is quite consistent with the results obtained by Metzger (1959) for paired-associate gating tasks with constant response uncertainty. These results indicate that the clear distinction between gating and condensation tasks is lost in concept-learning studies. Both are increasing linear functions of the amount of information reduced. This is reasonable since the *S* must learn what aspects of the stimulus are relevant and as the stimuli are made more complex this increases the difficulty. Once the gating rules are learned, however, the gating tasks come to resemble conservation tasks and little effect of irrelevant information on processing time is found (Archer, 1954).

In the concept-identification experiments cited above the number of relevant stimulus levels always equals the number of responses. Battig and Bourne (1961) have discussed this limitation as follows, "one shortcoming of the concept-identification task, and of many other procedures used in experiments purportedly concerned with concept formation, is that in a real sense they do not require concepts to be formed. This is because all of the stimuli to be classified with the same response are identical with respect to the crucial or relevant characteristics [p. 329]."

Condensation. Battig and Bourne (1961) have attempted to remedy this defect by varying both the number of irrelevant dimensions and the variability within a response category. In order to do this they compared, for example, a situation where a single response was made to right, equilateral, and obtuse triangles and another to squares, trapezoids, and parallelograms. In this situation a very simple recoding (four sides versus three sides) could allow *S* to ignore some of those differences. To determine the actual extent of such recoding the stimulus patterns should be tried in a reaction time task to find out if after learning the RT would approach that for a one bit conservation task. In any case, regardless of the ability of *S* to ignore such differences, the results so far reviewed in the learning of concepts (Archer, Bourne, & Brown, 1955; Metzger, 1959) would indicate that a linear relation should exist between the amount of information reduced in forming the response class and rate of learning. Hunt (1962) analyzed the Battig and Bourne data and found that the relation between errors to criterion and information within a response category is linear.

The concept-learning study most similar to the condensation tasks discussed in section one is that of Shepard, Hovland, and Jenkins (1962). In these experiments there were always eight stimulus patterns, consisting of two levels on each of three dimensions. There were always two response categories. The problems differed in the way in which the information relevant to the classification was distributed across the dimensions. In one problem all the information involved only one dimension, in another all information involved two dimensions simultaneously, a third type had all three dimensions relevant, and a final type involved some of the information in one, two, and three dimensional terms. The total amount of relevant information within a response category could be calculated by weighting the proportion of information by the number of dimensions which had to be attended in order for it to be extracted. By this method the four types of problems had 1, 2, 2.3, and 3 bits of relevant information, respectively. Figure 6 shows the success of concept learning as a function of the amount of information condensed in obtaining the binary classification. The points of this figure are from Tables 4 and 6 of the original report and represent averages over different sets of a given problem type and over different display conditions. This result indicates that condensation is more important than gating when they are opposed, as in this task. There is some indication that with more extensive training than that shown in Figure 6 the 2-bit reduction tasks begin to become easier than some of the 1.3-bit tasks.

In the learning of concepts the amount of information reduced appears to be a useful summarizing variable. As well as having a consistent relation to task difficulty it also demonstrates the close relation between unidimensional and multivariate concept learning both conceptually and empirically. This measure and the taxonomy based upon type of information processing should be of significant value in an integrated attack upon complex information processing.

REFERENCES

Archer, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

Archer, E. J., Bourne, L. E., Jr., & Brown, F. G. Concept formation as a function of irrelevant information and instruction. *J. exp. Psychol.*, 1955, **49**, 153-164.

Attneave, F. Perception and related areas. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 4. New York: McGraw-Hill, 1963.

Bartlett, F. C. *Thinking: An experimental and social study.* New York: Basic Books, 1959.

Battig, W. F., & Bourne, L. E., Jr. Concept identification as a function of intra- and interdimensional variability. *J. exp. Psychol.*, 1961, **61**, 329–333.

Bourne, L. E., Jr., & Haygood, R. C. The role of stimulus redundancy in concept identification. *J. exp. Psychol.*, 1959, **58**, 232-238.

Bricker, P. D. Information measurement and choice time: A review. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955. (a)

Bricker, P. D. The identification of redundant stimulus patterns. *J. exp. Psychol.*, 1955, **49**, 73–81. (b)

Broadbent, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

Bruner, J. Going beyond the information given. In, *Contemporary approaches to cognition.* Cambridge, Mass.: Harvard Univer. Press, 1957.

Campbell, D. T. Systematic error on the part of human links in communication systems. *Inform. Cont.*, 1958, **1**, 334–369.

Crossman, E. R. F. W. Entropy and choice time: The effect of frequency imbalance on choice-response. *Quart. J. exp. Psychol.*, 1953, **5**, 41–51.

Fitts, P. M. Human information handling in speeded tasks. IBM Research Report 109, 1959.

Fitts, P. M., & Biederman, I. S-R compatibility and information reduction. *J. exp. Psychol.*, 1965, **69**, in press.

Fitts, P. M., Peterson, J. R., & Wolfe, G. Cognitive aspects of information processing: II. Adjustments to stimulus redundancy. *J. exp. Psychol.*, 1963, **65**, 423–432.

French, R. S. Number of common elements and consistency of reinforcement in a discrimination learning task. *J. exp. Psychol.*, 1953, **45**, 25–33.

Garner, W. R. Uncertainty and structure as psychological concepts. New York: Wiley, 1962.

Gerard, R. W. Neuro-physiology: An integration. In, *Handbook of physiology*, Vol. 3. Washington, D. C.: American Physiological Society, 1960.

GREGG, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

HAKE, H. W., & ERIKSEN, C. W. Effect of number of permissable response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.* 1955, **50**, 161–167.

HUNT, E. B. *Concept learning: An information processing approach.* New York: Wiley, 1962.

METZGER, R. A comparison between rote learning and concept formation. *J. exp. Psychol.*, 1958, **56**, 226–231.

MILLER, G. The magical number seven; plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

MORIN, R. E., FORRIN, B., & ARCHER, W. Information processing behavior: The role of irrelevant stimulus information. *J. exp. Psychol.*, 1961, **61**, 89–96.

MORIN, R. E., & FORRIN, B. Response equivocation and reaction time. *J. exp. Psychol.*, 1963, **66**, 30–36.

PIERCE, J. R., & KARLIN, J. E. Reading rates and the information rate of a human channel. *Bell Sys. tech. J.*, 1957, **36**, 497–516.

POLLACK, I. Speed of classification of words into super-ordinate categories. *J. verbal Learn. verbal Behav.*, 1963, **2**, 159–165.

POSNER, M. I. An informational approach to thinking. Technical Report, Office of Research Administration Project 02814, Office of Research Administration, Ann Arbor, April 1962.

POSNER, M. I. Uncertainty as a predictor of similarity in the study of generalization. *J. exp. Psychol.*, 1964, **68**, 113–118.

RABBITT, P. M. Ignoring irrelevant information. *Amer. Psychologist*, 1963, **18**, 472. (Abstract)

RICHARDSON, J. The relationship of stimulus similarity and number of responses. *J. exp. Psychol.*, 1958, **56**, 478–484.

SHEPARD, R. N., & CHANG, J. J. Stimulus generalization in the learning of classification. *J. exp. Psychol.*, 1963, **65**, 94–102.

SHEPARD, R. N., HOVLAND, C. J., & JENKINS, H. M. Learning and memorization of classifications. *Psychol. Monogr.*, 1961, **75**(13, Whole No. 517).

SMITH, T. A., JONES, L. V., & THOMAS, S. Effects upon verbal learning of stimulus similarity, number of stimuli per response and concept formation. *J. verbal Learn. verbal Behav.*, 1963, **1**, 470–476.

SPERLING, G. A model for visual tasks. *Hum. Factors*, 1963, **5**, 19–31.

TODA, M. Review of, *Uncertainty and structure as psychological concepts. Psychometrika*, 1963, **28**, 293–310.

WALKER, C. N., & BOURNE, L. E., JR. The identification of concepts as a function of amounts of relevant and irrelevant information. *Amer. J. Psychol.*, 1961, **74**, 410–415.

WERTHEIMER, M. *Productive thinking.* New York: Harper, 1959.

(Received October 21, 1963)

Psychological Review
1964, Vol. 71, No. 6, 505–513

THEORETICAL NOTES

A POWER COMPARISON OF THE *F* AND *L* TESTS—I

FREDERICK J. BOERSMA, JAMES J. DEJONGE,[1] AND
WALTER R. STELLWAGEN

Michigan State University

The report deals with the behavior of the omnibus F test in a restricted hypothesis space and attempts to answer the question "Is it possible that a nonparametric test may be considerably more powerful than the F test even under conditions of strict normality and equality of variance?" An empirical power comparison was made between the F and L tests by generating populations having specified ordered relationships and equal variances for various combinations of treatments and replications. The percentage of rejections of the null hypothesis was plotted for both tests for each set of parameters. The power comparison shows the nonparametric L test to be more powerful than the omnibus F test when there is a monotonic order among the treatment means. In conclusion, it is suggested that when theory and/or a priori considerations indicate ordered treatment affects, the experimenter consider the use of statistical techniques which deal with restricted hypothesis spaces.

The testing of ordered hypotheses is of special interest to behavioral scientists because many theories and a prior considerations suggest ordered treatment effects. In spite of this, few statistical tests are available for testing such hypotheses. Many tests currently used are essentially generalizations of tests for evaluating the significance of differences between treatment means.

Probably the only adequate two-sample parametric test which is available for evaluating ordered hypotheses is the one-tailed t test. This test is quite powerful when the assumptions of normality and homogeneity of variance are met and reasonably powerful when these assumptions are violated as long as the variances are equally weighted and the distribution is not too extreme.

There are also a number of two-sample one-sided nonparametric tests available for testing ordered hypotheses. Among these are the binomial test, the sign test, the Mann-Whitney U test, and the Kolmogorov-Smirnov two-sample test (Siegel, 1956).

It is generally assumed that the one-tailed t test is more powerful than its nonparametric counterparts when the underlying assumptions are met.[2] It has been shown, however, that when sampling is from nonnormal distributions, the U test tends to be as powerful as the t test (Boneau, 1962).

For more than two treatments, the problem of testing directional hypotheses becomes somewhat more complicated. The omnibus F test is generally used in such situations since it is considered to be the most powerful test available for testing for differences between means. It should be noted, however, that an omnibus test is inappropriate for testing ordered hypotheses since it is affected by mean differences in any direction and will reject H_o with high probability even when the ordered hypothesis is false. Either because of the assumption of best power or because of a certain naiveté concerning the theory of statistical tests, which expresses itself most clearly in the phrase "one- versus two-tailed tests," a practice has arisen in the psychological literature of performing a directional statistical test only if an omnibus test has already led to the rejection of the null hypothesis. This paper will only concern itself with the problem of power. It might be well to examine, however, the issue of tests of directional versus nondirectional statistical hypotheses.

---

[1] Now at the United States Army Personnel Research Office, Washington, D. C. 20540.

[2] That the t test is not uniformly most powerful at the usual significance levels has been shown by Lehmann and Stein (1948).

Under the restriction $\sum_{i=1}^{K} \mu_i = 0$, the hypothesis space of the omnibus F test is a $k - 1$ dimensional space. The null hypothesis consists of the origin of the space and the alternative hypotheses consist of the remainder of the space. Because the omnibus F test has best average power on the surface of Euclidean spheres about the origin of this space, it is probably the most appropriate test to use when all possible points in the alternative hypothesis space are of equal interest. However, it is rarely the case that all points are equally of interest; a few large differences are usually more interesting than many small ones, direction is usually of some importance, etc. Therefore, various nonomnibus tests have been devised. These tests will have better power than the omnibus F test in certain regions of the hypothesis space and lesser power in other regions.

In psychological literature, arguments concerning directional tests are most commonly phrased in terms of "significance level." The use of a restricted test is felt to (spuriously) "increase the significance" of any given result. In the case of the t test, the argument seems almost reasonable. The critical region for the one-tailed t test contains all the points in the critical region of the two-tailed t test for which the sample mean is in the hypothesized direction. We must note, however, that there are samples for which the two-tailed test will reject H_o and the one-tailed will not. In the general case, the argument loses all its force. The restricted tests have critical regions of entirely different shapes from the omnibus tests. The choice of a test should be made on the basis of its power for the alternative of interest. Tests should be compared only with respect to the same hypothesis space. Only tests with equally sized critical regions should be compared.

The question left then is that of power. Is there a test which is more powerful than the omnibus F test when the hypothesis space is restricted? The answer is obviously yes. However, it might be of some interest to demonstrate that *even under conditions of strict normality and equality of variance* a nonparametric test may be considerably more powerful than the omnibus F test. This is the problem which will be considered in this paper. Attention should first be directed, however, to statistical techniques available for handling k treatments.

The orthogonal polynomial analysis for the comparison of curves in the analysis of variance is a parametric approach to this problem. Grant (1956) presents statistical tests for various polynomial components of the treatment sum of squares and advocates the use of such tests when the assumptions of independence, normality, and equal intervals between the levels of the independent variable are met. This regression approach allows one to determine whether there exists an overall trend. Edwards (1960), McNemar (1962), and Winer (1962) explain this approach further.

Bartholomew (1961) has developed a technique for testing ordered hypotheses without requiring the assumption of equal intervals between the independent variables. Bartholomew's test is superior to the omnibus F test when a priori considerations suggest ordered hypotheses and appears to be the best multiple treatment parametric test available for testing such hypotheses.

Nonparametric methods which make use of the one-way analysis of variance approach are given by Armitage (1955) for nominal data and by Jonckheere (1954a) for ordinal data.

Probably the most widely used nonparametric technique for situations involving more than two treatments is Friedman's (1937) two-way analysis of variance by ranks for related samples. This chi-square test assumes only an ordinal scale of measurement and tests whether the samples could have come from the same population. Friedman (1940) finds this test to be as powerful as the omnibus F test when the assumptions and requirements of the latter are met.

Jonckheere's (1954b) two-way analysis of variance test is based upon Kendall's "average tau." It is the difference between a predicted theoretical ranking and a number of judges' rankings. Another ranking statistic is Page's L test (1963) which is specifically designed for testing ordered hypotheses in the randomized blocks design analysis of variance. It is computationally much easier than Jonckheere's test and is favored over its parametric counterpart in that it requires only an ordinal scale of measurement.

The L test is closely related to Friedman's χ_r^2, the randomized blocks design analysis of variance, regression analysis, and correlation. Page shows this relationship and gives tables associated with the use of his test. He suggests use of the L test as a replacement for Friedman's test. A limited power comparison made by him shows L to be more powerful than Friedman's test for testing ordered hypotheses. Since Friedman found a negligible power difference between his test and the omnibus F, we may assume that the L test will be at least as powerful as the omnibus F test.

The purpose of the present study is to show empirically that the nonparametric L test is more powerful than the omnibus F test when a priori hypotheses suggest ordered relationships, even though the assumptions of normality and homoscedasticity have been met. A secondary problem which will be considered in the future is a power comparison of the L test with an F test for linear trend.

As noted above, the most frequent practice is to perform the omnibus test first and follow with a directional test only if the omnibus test leads to a rejection of H_o. We do not maintain that this is a good practice, but it is almost universal in application. We note, therefore, that *the power of the method most commonly in current use cannot exceed that of the omnibus* F *test*. A comparison of L with a parametric analog may be of considerable interest. If the means are ordered and assumptions of normality and homoscedasticity are met, the question is raised, "How does the power of the omnibus F test compare with that of L?" If power differences are negligible we might expect that L might be more powerful under conditions of nonnormality and heteroscedasticity.

METHOD

The procedure used in the present research involved calculating a large number of F and L values, each based upon samples which were drawn at random from distributions having specified characteristics, and constructing the power functions of the obtained values. The study was performed on Control Data Corporation 160 and 3600 electronic computers programed to perform the following operations:

1. Generation of a uniformly distributed random number
2. Transformation of a sequence of these numbers into a random deviate from a normal distribution
3. Successive generation of random deviates until a sample of certain size is reached and transformation of these deviates by adding mean values
4. Calculation of the F and L value based on the sample
5. Calculation of the power of the F and L tests on the basis of a series of such samples

The complete sequence of operations was performed within the computer. The end result, the analysis of variance tables, L values, and the power of the respective tests, were printed by the computer.

Amplification of the above operations may be relevant and will be made in the order just given.

1. The random number consisted of 8 digits, each being equal to the sum of the previous two random numbers mod 10^8. To start the process it was necessary to place in the machine two 8-digit random numbers selected from a table of random numbers.

2. In order to obtain a random normal deviate, a sequence of 10 uniformly distributed random numbers obtained in the

above fashion was added[3] to produce a single random deviate having zero mean and unit variance. The medians, means, and variances of the deviates generated in this way were tested by the usual tests and found to be within the limits of chance.

3. The number of replications per treatment (N) selected in the present investigations was 5, 10, and 15. The number of treatments (k) was 3, 4, and 5.

The major problem in computing power in a situation involving k groups is the selection of mean values, μ_i. Without loss of generality, assume $\sum_{i=1}^{k} = \mu_i = 0$ and $\sigma^2 = 1$. The hypothesis space in this experiment is then the origin plus the smaller Region R of the k–1 dimensional space bounded by the hyperplanes $\mu_i = \mu_j,\ i \neq j$. A line passing through the origin represents the condition that the μ_i be in a specified ratio. The distance of a point from the origin is the noncentrality parameter η, utilized in computing the power of the F test. While the power of the omnibus F test depends only upon η, the power of other tests may depend on the μ_i themselves. Among the class of tests which depend only on η, the F is known to be uniformly most powerful.[4] Also, F is a test with maximum average power on such spheres. However, other tests probably have better average power on other surfaces. Short of a point by point comparison, there appears to be no way to offer a contrast fair to both tests. It was, therefore, decided to select arbitrarily a line passing through the origin, to select points along this line, and to plot the power against η. This is, at least, fair to F since for fixed η the power of F is uniform. The line selected was the line defined by the points $\mu_1=0$; $\mu_{i+1}=\Delta(\mu_i+1)$. Points on the line were selected by multiplying the coordinates of the point $(0, 1, 2 \cdots k)$ by a scalar Δ.

4. The formula used for the calculation of L was the following (Page, 1963):

$$L = \sum_{j}^{k} (Y_j \sum_{i}^{n} X_{ij})$$

where Y_j is the hypothetical ranking of the jth treatment and $\sum_{i}^{n} X_{ij}$ equals the sum of ranks observed in the jth column. F was calculated from the appropriate formula for the randomized blocks design analysis of variance.

5. For every combination of number of treatments, replications, and mean differences between treatment groups, 25 samples were generated. The power of each test was given by the proportion of times it rejected the null hypothesis.

To summarize, random samples were drawn from populations which were normally distributed but had different means. For several combinations of treatments and sample sizes, F and L values were calculated and the power of each test computed as a function of the percentage of rejections of the null hypothesis.

RESULTS

Samples were generated for the various combinations of treatments and sample sizes. The percentage of times that the null hypothesis was rejected by F and by L was calculated for each combination at three significance levels. Running averages of these power functions were plotted against the logarithm (4 cycles by 10 to the inch) of the noncentrality. The difference between the L and F curves for a particular alpha level empirically shows the power difference between the two tests. For an explanation of power and noncentrality, see Scheffé (1959, pp. 38–39).

Looking at all curves we note the power differences are greater in the middle than at the ends. This is to be expected, however, since all power curves must be at height alpha at H_o and most curves are at height 1.0 as noncentrality increases to large values.[5]

The first case to be considered is that of three treatments. Figure 1 shows the power curves associated with five replications per treatment. L is shown to be considerably more powerful that F[6] at the .05 and .01 levels. This is especially true for noncentrality values between .50

---

[3] A sequence of random deviates (X_j) is given by $X_j = (\sum_{i=10(j-1)}^{10j-1} R_i - 5.0)\ (\sqrt{1.2})$, where $R_{i-1} = mR(\text{mod } 10^8)$ and m is some prime number.

[4] Scheffé (1959, p. 48). In the general case η would be $\frac{n \sum_i (\mu_i - \mu_0)^2}{\sigma^2}$; in the present case this reduces to $\Sigma\mu_i^2$, since $\sigma = 1$ and n and $\mu.$ are fixed.

[5] Zero noncentrality was not run since $\log 0 = -\infty$.

[6] Unless otherwise indicated hereafter "F" refers to the omnibus F test.

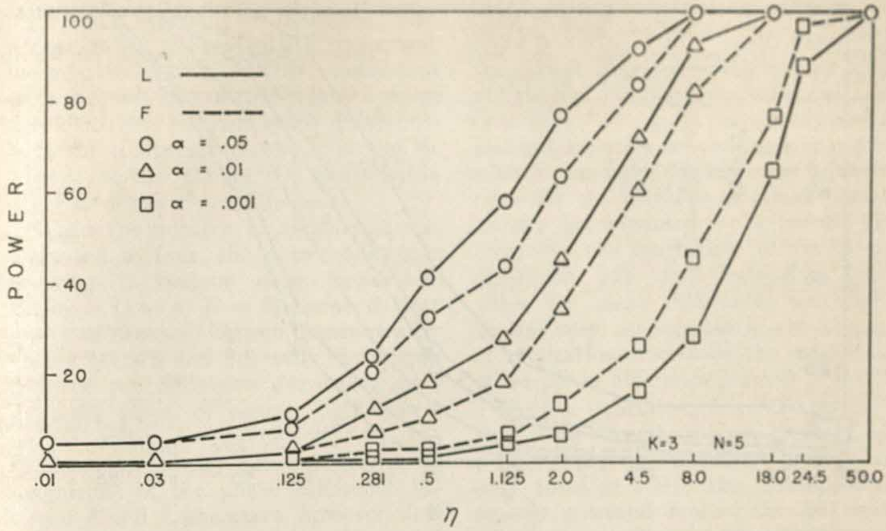


FIG. 1. Power curves for the F and L tests plotted against η when $k = 3$ and $N = 5$.

and 4.50. At the .001 level L was found to be consistently less powerful then F, especially between noncentrality values of 2.00 and 24.5.

When the number of replications per treatment is 10, there is a very small power difference in favor of the L test at all alpha levels. The greatest power differences in favor of L were found at the .05 and .01 levels with small noncentrality values. The above tends to indicate that when a priori considerations suggest small monotonic differences between three treatment effects, L would be preferred over F at the .05 and .01 levels. There is no apparent power

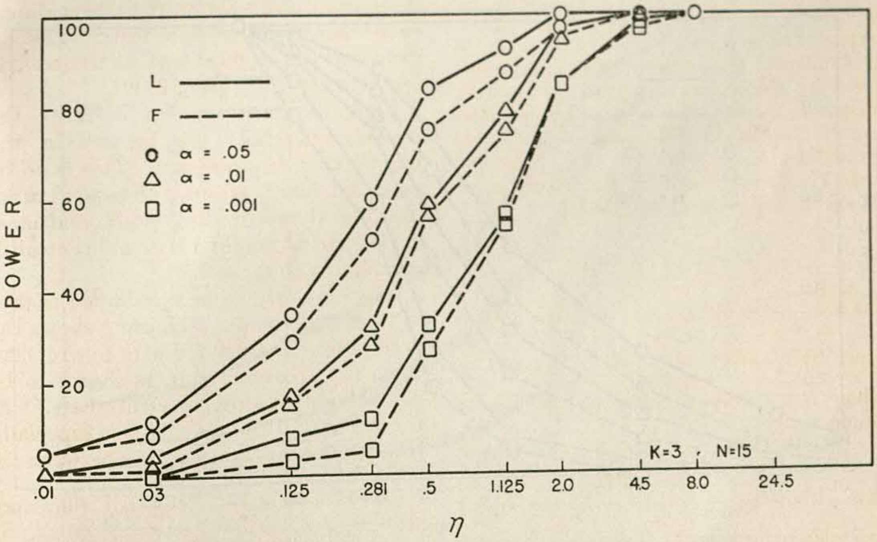


FIG. 2. Power curves for the F and L tests plotted against η when $k = 3$ and $N = 15$.

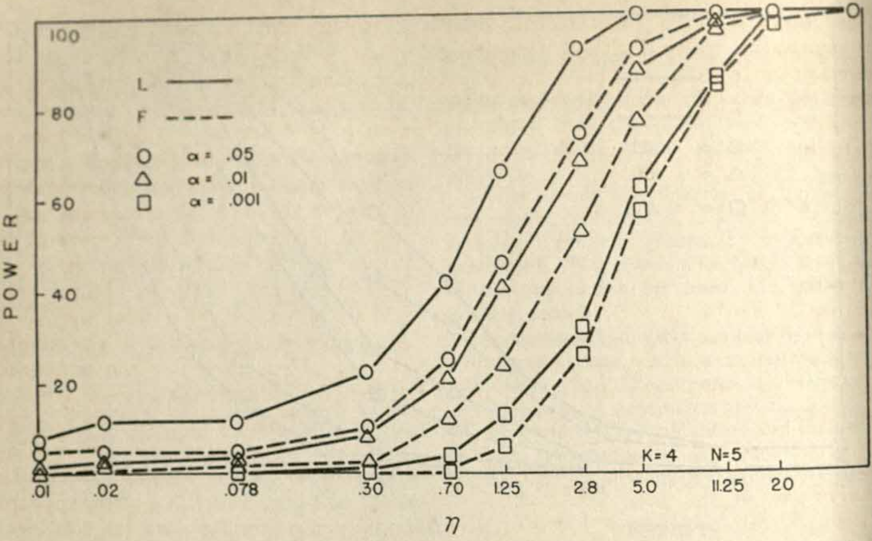


FIG. 3. Power curves for the F and L tests plotted against η when $k = 4$ and $N = 5$.

difference between the L and F tests at the .001 level with 10 replications.

When the number of replications equals 15 (see Fig. 2), L tends to be consistently more powerful than F at the .05 level. This is particularly evident for noncentrality values between .03 and 2.0. The power difference at the .01 and .001 levels also favor L. However, very little power can be gained by the use of L at these levels.

These results indicate that L will be considerably more powerful than F when the number of replications is small for noncentrality values greater than .125. The power difference below this point appears negligible. When the number of replications increases both tests be-

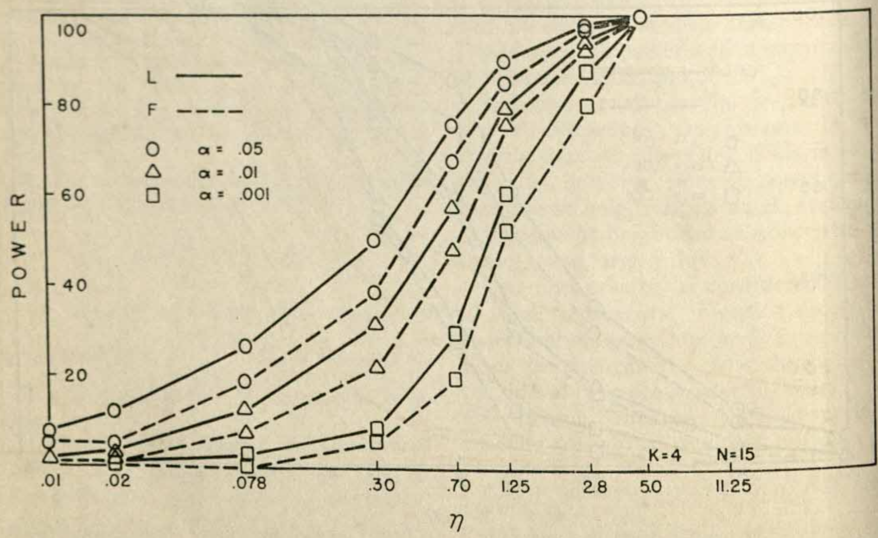


FIG. 4. Power curves for the F and L tests plotted against η when $k = 4$ and $N = 15$.

come more powerful for detecting small noncentrality values with L remaining more powerful than F. This is consistent with the well-known rule that if a priori considerations indicate small differences between treatment means it would be wise to increase sample size somewhat in order to detect these differences.

When the number of treatments was increased to four, the power difference favoring L became more noticeable. Figure 3 ($k = 4$, $N = 5$) shows a large power difference betwen noncentrality values of .078 and 5.0 with the magnitude of the difference decreasing as η increases since, of course, both power curves must be asymptotic to 1.0. When noncentrality falls below .078 the magnitude of the power difference between F and L decreases, however, L is still the most powerful. Figure 4 ($k = 4$, $N = 15$) shows power curves similar to Figure 3, the only exception being that L shows a greater power difference at the .05 and .01 levels for small mean differences. At the .001 level, when noncentrality was equal to or greater than .078, the power difference favoring L was much greater than in the previous case. This is consistent with the findings for three treatments.

In summary, for four treatments L is more powerful than F in all cases with the maximum power differences occurring at low and intermediate values of noncentrality. For large values of noncentrality both tests are extremely powerful with L remaining slightly more powerful than F. An increase in sample size is usually accompanied by a general decrease in the magnitude of the power difference, the only exception being when the mean differences are small. In this case, an increase in the number of replications increases the magnitude of the power difference between L and F.

For five treatments L is more powerful than F for all sample sizes, levels of significance, and noncentrality values. The only point at which the two appeared equally powerful was at the .001 level when noncentrality is 5.60 and greater.

Figures 5 and 6 show power curves consistent with previous figures. When the number of treatments is 5 and the number of replications is 5, 10, and 15, large power differences favoring L were found at the .05 and .01 significance levels. This power difference was especially noticeable at the .05 level for small noncentrality values. Large power

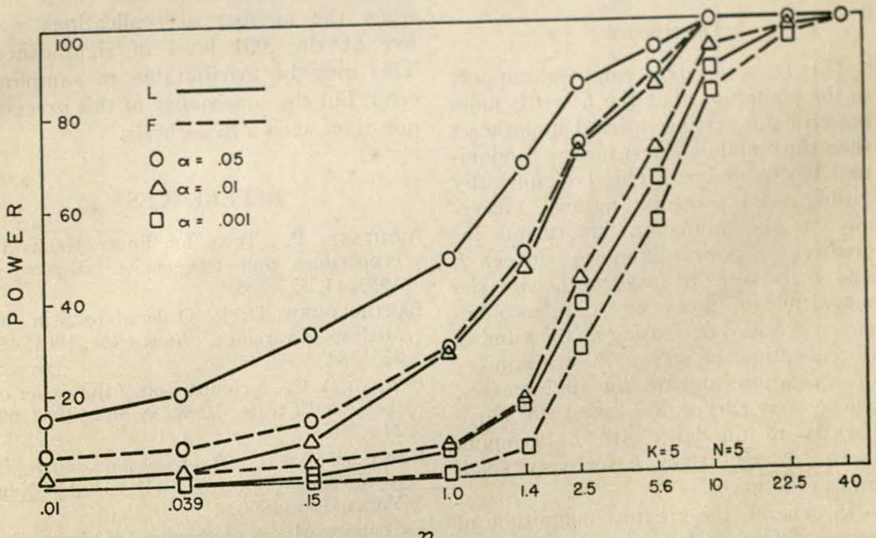


FIG. 5. Power curves for the F and L tests plotted against η when $k = 5$ and $N = 5$.

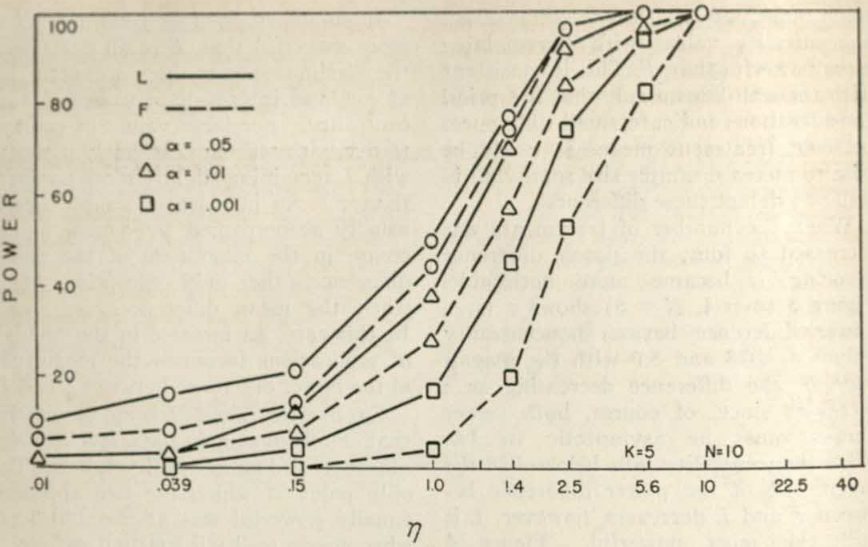


FIG. 6. Power curves for the F and L tests plotted against η when $k = 5$ and $N = 10$.

differences were not found, however, for small mean differences.

In conclusion, if hypotheses suggest large differences between treatment means, increasing sample size will increase the power of both tests with L remaining about 10 to 15% more powerful.

DISCUSSION

This research gives empirical support to the prediction that the L test is more powerful for testing ordered hypotheses than the omnibus F test for the randomized blocks design. This is empirically visible in the preceding figures. Generally, as the number of treatments increases, the power difference between L and F increases in favor of L with the magnitude of the power difference between F and L decreasing as the number of replications increases. As the number of replications and treatments increases, the power curves also tend to move slightly to the left, with L becoming more powerful than F for very small mean differences.

In general, the greatest magnitude of power difference between the L and F tests appears to be at the .05 and .01 levels. This is most noticeable when the number of replications is small. There is an increase in power for both tests as noncentrality increases with L being consistently more powerful.

The only case where F was found to be more powerful than L was when the number of treatments was three and when the number of replications was five at the .001 level of significance. This may be attributable to sampling error, but the consistency of this reversal questions such a hypothesis.

REFERENCES

ARMITAGE, P. Tests for linear trends in proportions and frequencies. *Biometrics*, 1955, **11**, 375–386.

BARTHOLOMEW, D. J. Ordered tests in the analysis of variance. *Biometrika*, 1961, **48**, 325–332.

BONEAU, A. C. A comparison of the power of the U and t tests. *Psychol. Bull.*, 1962, **69**, 246–256.

EDWARDS, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Holt, 1960.

FRIEDMAN, M. A comparison of alternative tests of significance for the test of *m* rankings. *Ann. math. Statist.*, 1940, **11**, 86–92.

FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.*, 1937, **32**, 675–701.

GRANT, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

JONCKHEERE, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145. (a)

JONCKHEERE, A. R. A test of significance for the relation between *n* rankings and *k* ranked categories. *Brit. J. statist. Psychol.*, 1954, **7**, 93–100. (b)

LEHMANN, E., & STEIN, C. Most powerful tests of composite hypotheses: I. Normal distribution. *Ann. math. Statist.*, 1948, **19**, 495–516.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1962.

PAGE, E. B. Ordered hypotheses for multiple treatments: A significance test for linear ranks. *J. Amer. Statist. Ass.*, 1963, **58**, 216–230.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received December 2, 1963)

Psychological Review
1964, Vol. 71, No. 6, 514–516

A COGNITIVE INTERPRETATION OF INTENSITY EFFECTS IN STIMULUS GENERALIZATION

FRANK RESTLE

Indiana University

A subject trained to a dim light responds well to a stronger light, but if trained to the stronger light he responds weakly to the dim. Hull calls this stimulus-intensity dynamism (V). If stimuli are represented as sets and a stronger stimulus as a larger set, then stimulus-sampling theory predicts an effect opposite to that usually observed. If learning amounts to selection of a strategy, as in cognitive theories, and if stronger stimuli lead to larger sets of adequate strategies, then weak-to-strong generalization is greater than strong-to-weak. This note formalizes and generalizes the cognitive theory, using a set-theoretic model.

Consider two stimuli of different intensity, letting S_5 be the more intense and S_1 the less intense. In an experiment, Group $G_{1\rightarrow5}$ is first trained to respond to Stimulus S_1, then is tested on S_5. Another group, Group $G_{5\rightarrow1}$, is first trained to respond to S_5 and is then tested on S_1. It is repeatedly found in such experiments that the probability of response to Stimulus S_5 by Group $G_{1\rightarrow5}$ is higher than the probability of response to S_1 by Group $G_{5\rightarrow1}$ (Mednick & Freedman, 1960).

Hullian Theory

This finding is discussed as Stimulus-Intensity Dynamism (V) by Hull (1949), who postulated that response strength is influenced by the intensity of the test stimulus. Since S_5 is more intense than S_1, it provides more dynamism. Since Group $G_{1\rightarrow5}$ is tested on the more dynamic stimulus, it responds strongly. Since Group $G_{5\rightarrow1}$ is tested on the weak stimulus, it responds weakly. True underlying generalization is assumed to be symmetrical.

The postulate of Stimulus-Intensity Dynamism has little independent support, and does not deduce the phenomenon from more elementary principles. A more fundamental analysis is difficult to obtain from Hullian theory, for in that theory there is no natural representation of stimulus intensity, as distinguished from qualitative variations of stimuli.

Stimulus-Sampling Theory

If stimulus situations are represented as sets, the intensity of a stimulus may be represented by the number or measure of elements in it. Recent discussions have shown that a dimension of stimuli can each be represented as a set, and the sets can be arranged in an orderly way (Restle, 1959, 1961a).

One theory of generalization (Bush & Mosteller, 1951) employs sets in the framework of stimulus-sampling theory. Let S and S' be two stimuli to which the subject initially does not respond. Suppose the subject is then trained to respond to Stimulus S with probability p. From various special assumptions it is deduced that the probability of a generalized response to S' is p', given by,

$$p' = p\,\frac{m(S \cap S')}{m(S')}, \qquad [1]$$

where m is a suitable measure function. Suppose that the training probabilities p are equal for Groups $G_{1\rightarrow5}$ and $G_{5\rightarrow1}$. By definition the intersections will be $m(S_1 \cap S_5)$ and $m(S_5 \cap S_1)$, which are equal. Hence, the probability of generalized response will be inversely proportional to $m(S')$, the measure of the test stimulus.

This means more generalization to S_1 in Group $G_{5\to 1}$, and less generalization to S_5 in Group $G_{1\to 5}$, the opposite of Mednick and Freedman's (1960) conclusion.

There are several possible retreats for the set-theory model. It could be argued that the experiments in question do not satisfy the important assumptions of "equal proportions" assumed by Bush and Mosteller (1951). However, systematic employment of the stimulus-sampling theory (Estes & Burke, 1953) seems to insure that the average of a group of subjects would satisfy the equal proportions assumptions, since training will have progressed by random sampling of stimulus elements from the Training Situation S. Another argument would be that it is naive to represent the more intense stimulus by a larger set. It can be argued that incidental effects could have been introduced into the experiments, or that the Training Level p was not controlled, or that the empirical results are not extremely impressive when looked at in detail.

Selection of Strategies: One at a Time

The writer (Restle, 1961b, 1962) has used ideas from older cognitive learning theories in a "selection of strategies" theory. Suppose that the subject acquires the original response by selecting some one strategy; that is, by selecting a particular aspect of the situation and an appropriate response. If the response has a satisfactory outcome, it is assumed that the subject continues to use exactly the same strategy.

Now imagine that the training situation presents the subject with a Set H of strategies of which a Subset R lead to the response. Let $m(H)$ and $m(R)$ be the measures of these sets. If a subject samples a single strategy from H, the probability it is in R is $m(R)/m(H)$.

In training the subject samples randomly from H. If his strategy leads to no response (is in $\bar{R}$) then the experimenter, by nonreinforcement or other device, induces the subject to resample. If at some point in training an element of R is chosen the experimenter rewards the response or otherwise induces the subject to keep that strategy. Let $P(R)$ be the probability that training has finally led the subject to select a strategy from R.

The test situation makes available a Set H' of strategies of which the Subset R' lead to response. From elementary probability theory,

$$P(R') = P(R'|R)P(R) + P(R'|\bar{R})P(\bar{R}). \quad [2]$$

Assume that $P(R'|\bar{R})$ is negligible. This idealizes the experiment by assuming that all responses to the test are generalized responses. Then,

$$P(R') = P(R'|R)P(R).$$

If the subject has a strategy in R he chose it in a single random sample. Therefore no subset of R is favored over any other, and,

$$P(R'|R) = \frac{m(R \cap R')}{m(R)}$$

whence,

$$P(R') = P(R)\frac{m(R \cap R')}{m(R)}$$

or in notation like that of Equation 1,

$$p' = p\frac{m(R \cap R')}{m(R)} \quad [3]$$

Equation 3 is similar to Equation 1. The differences are that Equation 3 refers not to the whole set but only to strategies which lead to response, and Equation 3 has, as the denominator of the fraction, the measure of the *training* situation not of the *test* situation.

By Equation 3, generalization is inversely proportional to the measure of the training set. This means more generalization to S_5 in Group $G_{1\to 5}$ and less generalization to S_1 in Group $G_{5\to 1}$. This is in agreement with the conclusions of Mednick and Freedman (1960).

Selection of Strategies: Random Sample

The results above depend upon the assumption that the subject uses exactly one strategy in R, but even if he should use more than one the same general kind of result follows. Suppose, for example, that the subject uses n strategies in R, and will generalize to R' if as many as one of these hypotheses appear in R'. Then the probability of generalization will depend upon the number of elements in R, R', and $R \cap R'$. Let the number of elements in R be r, the number in R' be r', and the number in the intersection $R \cap R'$ be i. Then assuming all elements have the same probability, the probability of generalizing is:

$$p' = 1 - \frac{\binom{r-i}{n}}{\binom{r}{n}}$$

$$= 1 - \frac{(r-i)!\,(r-n)!}{(r-i-n)!\,r!}$$

For fixed n and i, p' decreases with r. To prove this, call p' a function of r, and note that $[1 - p'(r+1)]/1 - p'(r)]$ is greater than 1.

Conclusions

The analysis given from cognitive theory has two advantages. Unlike Hullian theory, it has a natural and intrinsic way of representing intensity of stimuli, and predicts asymmetric generalization without any special postulate. Unlike the stimulus-sampling theory, which it otherwise resembles, the cognitive theory predicts the asymmetry in the direction observed in experiments.

REFERENCES

BUSH, R. R., & MOSTELLER, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.

ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

HULL, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.

MEDNICK, S. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57**, 169–200.

RESTLE, F. A metric and an ordering on sets. *Psychometrika*, 1959, **24**, 207–220.

RESTLE, F. *Psychology of judgment and choice.* New York: Wiley, 1961. (a)

RESTLE, F. Statistical methods for a theory of cue learning. *Psychometrika*, 1961, **26**, 291–306. (b)

RESTLE, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, **69**, 329–343.

(Received December 11, 1963)

Psychological Review
1964, Vol. 71, No. 6, 517–523

ACTUARIAL METHODS AS APPROPRIATE STRATEGY FOR THE VALIDATION OF DIAGNOSTIC TESTS[1]

JACOB O. SINES

University of Missouri School of Medicine

The equivocal results of research conducted to validate the clinical use of psychological tests are attributed to inappropriate design or strategy. The point is made that the predictions of behavior on the basis of test performance in the clinic does not necessarily imply the predictions which are examined in many validation studies. It is suggested that appropriate validation studies involve: (a) the evaluation of the hypotheses made by the clinician, (b) about test patterns rather than isolated aspects of the entire test responses. Such an approach seems to epitomize the use of actuaarial techniques. Data are provided to illustrate several of the critical considerations.

The published results of studies of the predictive and concurrent validity of most psychological tests do not seem to have seriously curtailed the clinical use of such techniques as the Minnesota Multiphasic Personality Inventory (MMPI), Rorschach, Thematic Apperception Test, or even in some quarters the Draw-a-Person Test (DAP). Many clinicians know such tests are useful while the skeptics know equally well that the conventional methods which have been used to assess the validity of these test instruments do not yield evidence to support this widespread faith.

This seeming inconsistency may be resolved in part if one closely considers the difference between the manner in which the clinician uses tests and the strategies most often used to study the validity of psychological tests. Brief attention to this difference may lead to a possible explanation of the equivocal results of predictive and concurrent validation studies and may at the same time suggest an alternative and possibly more productive validation procedure.

As the clinician uses the Rorschach or the MMPI he infers that persons who perceive the blots in a specific fashion or who generate a particular type of profile show certain characteristic behavior or dynamics. If one lets P_1 stand for a particular test pattern and B_1 refer to a specified pattern of real life behavior, the clinician's activity when describing personality on the basis of psychological test results may be represented logically as (if P_1 then B_1). As discussed by Hoffman (1960) this test-pattern factor (P_1) in the clinical judgment process seems to involve a number of test variables $X_1, X_2, \ldots, X_K$, each of which may be assigned selective weights, or the interaction of which may be considered by the clinician. Although not the main consideration at this time, it should be kept in mind that P_1, the test pattern, may, and usually does consist of several variables rather than just one variable or score. This point will be returned to below.

The skilled clinician bases his inferences on his past experience with persons who have responded to the test in more or less similar ways. The fact cannot be overemphasized that the clinician does not often or successfully predict a test pattern from a diagnosis, or even from several bits of socially relevant real-life behavior. It is indeed rare to find a clinician who can predict in great detail a patient's MMPI profile or Rorschach psychogram on the basis of brief contact with him. The frequency with which one encounters psychologists who can describe in great detail the behavior, thought pattern, or dynamics of some persons from only an MMPI profile or a Rorschach protocol is not unimpressive, however, and hopefully disturbing to the critics of psychological testing.

---

[1] The work reported here was supported in part by United States Public Health Service Grant MH 6097.

Unfortunately for the validity box score, many studies of the validity of psychological tests are not directly relevant to the question they were designed to consider. In most validation studies, a criterion group of persons is customarily tested and essentially the following prediction is put to test: "if a person shows certain real-life behavior then he will also show certain test behavior." Stated in the terms mentioned above the logical form of this statement is "if B_1 then P_1." Unfortunately this is quite a different matter from studying the validity claimed for the test by the clinician. Quite simply, to paraphrase Ainsworth (1954), the study of the validity of a test should involve an evaluation of the predictions made by the clinician when he uses that test. To repeat, the clinician's prediction is: "if a person shows a particular pattern of test behavior then he will show certain other relevant behavioral characteristics." Validity studies in which a discrete aspect of real-life behavior (suicide or response to therapy) is employed as the criterion variable, while regularity in a pattern of test behavior is viewed as the dependent variable may thus not only be irrelevant to the validity of psychological tests as they are used, but they seem to imply a widespread but unjustified notion of the logic of predictive relationships and a disregard for the possibility that multiple causes may be operating to produce a given behavior pattern.

The turmoil concerning the actuarial use of tests seems to have obscured the implications of one point stated by both Meehl (1954) and Holt (1958) which implies an alternative approach to the validation of psychological tests. In essence the actuarial approach as described by Meehl and the sophisticated clinical use of tests advocated by Holt, both involve the determination of the characteristic behavior, thinking pattern, or dynamics found among persons who respond to tests in certain common ways, that is, whose patterns of test performance are similar.

It must be admitted that the relationships between test variables and personality characteristics which Meehl might wish to formalize and which Holt considers important for the sophisticated clinician are not often clearly specified in the clinician's apperceptive mass, nor are they generally reflected in studies of the validity of psychological tests.

If one considers the assumed relationship between a socially significant behavior pattern, B_1, that one wishes to predict and various personality characteristics which may be reflected in Test Patterns P_1, P_2, or P_3, the above point may become clearer. On the basis of training and experience the clinician may have found on many occasions that Behavior Pattern B_1 is observed in persons who show Test Pattern P_1. This is not a logically symmetrical statement in that it in no way implies that Behavior Pattern B_1 may not also be found in persons who generate Test Patterns P_2 or P_3, or otherwise stated, it does not imply that every person who shows Behavior Pattern B_1 will also show Test Pattern P_1.

The usual study of predictive or concurrent validity is designed *as if* the clinician's behavior were "$P_1 \equiv B_1$," that is, "Behavior Pattern B_1 implies Test Pattern P_1 and Test Pattern P_1 implies Behavior Pattern B_1." Test Patterns P_2 and P_3 are too often not considered.

Thus, many of the studies of predictive or concurrent validity which offer only equivocal support for this latter hypothesis "if Behavior Pattern B_1 then Test Pattern P_1," should at least raise the possibility that several personality patterns exist which may be reflected in Test Patterns P_2, P_3, or P_4 as well as P_1, each of which may be predictively related to the behavior pattern we are interested in.

This more realistic conception of predictive relationships would lead to the hypothesis that "if P_1 (or P_2 or P_3) then B_1"; and this may in fact be more representative of the way the clinician functions in his actual work.

Consider the possibility that a given behavior pattern (B_1) may be predicted from any of five test patterns. For the sake of illustration let us assume that of

100 persons who manifest Behavior Pattern B_1, 40 generate Test Pattern P_1, and 15 generate each of the Test Patterns P_2, P_3, P_4, and P_5. The chances are good that the clinician's experience will lead him to state "if P_1 then B_1." If he should attempt to validate his conclusion by determining the average test pattern of a sample of persons who show Behavior Pattern B_1, he will be unsuccessful to the extent that he has a representative sample of persons who show Behavior Pattern B_1. His initial clinical observation (if P_1 then B_1) may have a very high degree of validity yet Test Pattern P_1 may account for less than half of the variance he finds in his attempt to validate that statement.

Gilberstadt (1962) has illustrated this point in a recent report in which he contrasts mean *and* modal MMPI profiles in a group of neurodermatitis patients.

It was noted above that there may be several variables and possibly their interaction which constitute the test-pattern factor in clinical judgment. This seems to be the essence of the frequently repeated caution that the diagnostic or predictive significance of a Rorschach $F+\%$ or an MMPI *Sc* score depends upon the entire test pattern in which it occurs.

Consider the implications if Behavior Pattern B_1 is in fact found in persons who generate some Test Pattern P_1. Test Pattern P_1 consists of several variables, $X_1, X_2, \ldots, X_K$, each of which are test representatives of personality factors which individually bear some predictive relationship to B_1, but no one of which alone explains the major portion of the variance in B_1.

We may study the predictive value of any one variable, X_j, in isolation if we wish, but logically we can expect a positive response to an inquiry in the form of "if X_j then B_1" in direct proportion to the weight of X_j in comparison to the summed weight of $X_1, X_2, \ldots, X_K$, or possibly some interaction of $X_1, X_2, \ldots, X_K$. Thus, even some studies of predictive or concurrent validity which appear to approximate the "if P_1 then B_1" form are actually tests of the hypothesis "if X_j then B_1."

A more reasonable and appropriate test of the predictive or concurrent validity of the Rorschach or the MMPI would seem to be "if personality (test) pattern P_1 then Behavior Pattern B_1" (if P_1 then B_1). And, furthermore, "if P_2 then B_1 and if P_3 then B_1." A meaningful test of an instrument's predictive or concurrent validity in these terms would necessarily involve the study of the real-life behavior of groups of persons whose *test behavior* is similar, rather than the study of the test behavior of diagnostic groups or groups of persons who manifest some limited amount of common real-life behavior. If persons whose test protocols or test patterns are highly similar also show highly similar social behavior, thought patterns, or dynamics, than the concurrent validity of a specific test pattern would be supported.

In order to test the above notions adequately, there appear to be at least three methodological considerations which must be clearly recognized, the last two of which were referred to by Holt (1950) some years ago.

First, and of critical importance, it must be kept in mind that the predictive relationship between a test pattern and a behavior pattern is asymmetrical. The statement "if B_1 then P_1" does *not* logically follow from the statement "if P_1 then B_1," the latter being the logical statement describing the clinician's use of his psychometric instruments.

Second, it is essential that we have a reliable method for the recognition of a test pattern. This requirement is most efficiently fulfilled at the present time if the test data are reducible to numerical form, as, for example, in the MMPI or the Holtzman Ink Blots. Several methods for the recognition of patterns or the calculation of indexes of profile similarity have been proposed (Meehl, 1959; Meehl & Dahlstrom, 1960; Mosel & Roberts, 1954). Most methods have been soundly criticized for their obvious deficiencies and while each has some degree of utility, no general or widespread application of any one of these methods to psychometric pattern recognition has appeared in the psychological literature to date. Suffice it to say that the investigator may choose from several methods, but whatever his choice, its weaknesses will be pointed out very quickly by his colleagues. In the study reported below the generalized distance function as described by Cronbach and Gleser (1953) was used to assess similarity of MMPI profiles.

Third and finally, one must determine the socially and psychologically relevant behavioral characteristics of the patients who generate a specific test pattern. It is on this point, too, that major disagreement may occur. The type of information one has about the persons already tested and the information one wishes to obtain by testing in the future will determine the source and type of personality data to be dealt with. There are those who rely primarily on various *Q* sorts (Marks & Seeman, 1963), terms used in psychiatric summaries (Gilberstadt & Duker, 1960), or statements derived from extensive clinical familiarity and previous research with a given test (Kleinmuntz, 1963; Pearson, 1962). None of the several sources of information are mutually exclusive and probably can be combined to provide the most comprehensive coverage. For a variety of reasons, the behavioral characteristics dealt with in the study reported below were drawn from the medical folders of psychiatric patients.

Illustrative Data

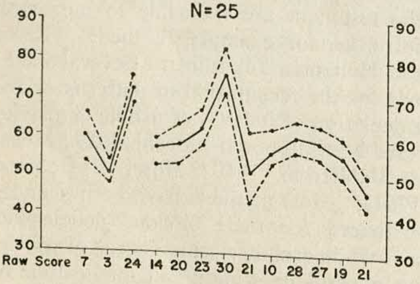


FIG. 1.

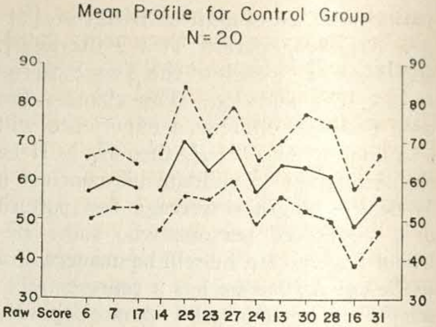


FIG. 2.

A group of 25 highly similar profiles was identified in a sample of 500 male MMPI profiles obtained at the time of admission to a state hospital.[2] Each profile in the sample was compared to every other profile and an index of profile similarity, the generalized distance function, *D*, was calculated for each pair of profiles (Cronbach & Gleser, 1950). The mean profile of this first cluster and the limits of ± 1 *SD* on each scale may be seen in Figure 1. A profile can be said to belong in this cluster (4′3 with *L* and *K* elevated) if it compares to the mean profile with a *D* of 22.00 or less ($D = \sqrt{\Sigma d^2}$, where *d* is the difference in *T* score points between corresponding scales of two profiles).

A control group of 20 profiles was selected from the same initial sample of 500 MMPI profiles. The control profiles were chosen if they did not compare with more than two other profiles with a *D* of 25.00 or less. The mean profile and the limits of ± 1 *SD* on each scale can be seen in Figure 2.

The hospital chart for each patient whose MMPI profile occurred in either the 4′3 group or the control group was then abstracted in detail. Each descriptive word, term, or phrase was recorded and sorted into one of 78 empirically derived categories.[3] In the following discussion only two of these categories, Personality Behavior and Personality Intrapsychic, will be considered.

---

[2] These data were obtained through the generous assistance of Howard R. Davis and William Leipold, Minnesota Department of Public Welfare.

[3] Other categories used in sorting descriptive statements in which relevant material may occur include: Affect, Cognitive Functioning, Interview Behavior, Mental Status, Relations with Parents, Social Behavior, as

The number of hospital records on patients in each of the above two groups in which a particular statement occurred was tallied and compared using a two-tailed test of significance. There were 125 different statements relating to "Personality" to be compared in these two groups of patients. Of these 125 comparisons, 18 were found to be significant at the .10 level or better; of these 14 were found at the .05 level or better and 4 at the .01 level or better. Those statements relating to Personality in our scheme which occurred in a significantly different proportion (at least the .10 level) of the medical records in the 4′3 group and the control group are listed in Table 1.

It can be seen that the incidence of assaultive-aggressive behavior is one of the outstanding characteristics (80%) of those patients who generated a 4′3 MMPI profile, while it occurs in only 40% of the control group of patients ($p < .01$).

In order to illustrate the asymmetry of the predictive relationship between the 4′3 MMPI pattern and the assaultive-aggressive behavior pattern a third group of patients was identified who had been admitted during the past year and were known by the hospital staff to have been assaultive or to have threatened assault. The period during which these patients were admitted overlapped for only a few months the admission period of patients whose profiles made up the original sample of 500 records. The medical records of this group were reviewed to verify the existence of assaultive behavior or threat of assault in every case. A group of 31 patients was located, each of whom had either threatened to kill or had seriously assaulted someone, and each of whom had generated an MMPI profile on admission to the hospital. The mean profile and the limits of ± 1 *SD* on each

TABLE 1

STATEMENTS IN PERSONALITY CATEGORIES FOUND SIGNIFICANTLY MORE FREQUENTLY OR SIGNIFICANTLY LESS FREQUENTLY IN THE 4′3 PROFILE GROUP THAN IN THE CONTROL GROUP (p = .10 or less)

Statement number	Frequency		Critical ratio	Statement
	4′3 Group (N = 25)	Control group (N = 20)		
0332058	20	3	4.334	Impulsive
0332081	20	8	2.949	History of hostile or aggressive behavior
0332009	7	0	2.575	Defensive
0333027	10	2	2.261	Lack of sense of responsibility
0332031	8	1	2.250	Seeks attention
0332064	8	1	2.250	Manipulative
0333043	7	1	2.121	Sensitive
0332059	5	0	2.121	Ingratiating manner
0333023	13	4	2.200	Immature
0332089	4	0	1.874	Said to be a follower
0333066	10	3	1.838	Uses projection
0332010	11	4	1.697	Demanding
0332098	0	6	2.941	Has a violent temper
0332035	1	6	2.391	Shows symptoms of anxiety and tension
0332102	0	3	2.004	Mild mannered
0332099	0	3	2.004	Felt to be dangerous
0332107	0	3	2.004	Has not threatened or injured others
0332048	2	6	1.918	Generous

---

well as others to be discussed in detail in a forthcoming report.

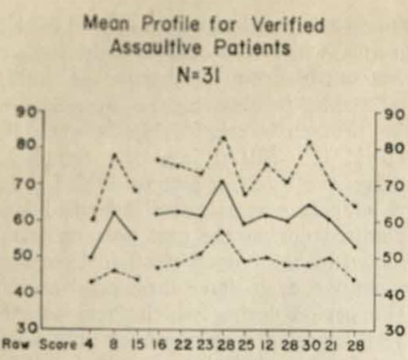


FIG. 3.

scale of the MMPI obtained from the assaultive patients are shown in Figure 3.

There was one patient in the "known assaultive" group who generated a 4′3 MMPI profile. The *D* value obtained from a comparison of the mean 4′3 profile and the mean profile of the assaultive group is 45.00, clearly larger than the *D* of 22.00 stated earlier as the defining parameter for inclusion in the 4′3 group.

On the basis of these data one can state with a reasonable degree of confidence that a 4′3 MMPI profile as defined here implies a behavior pattern which includes socially serious threats of assault or actual assault. The reverse does not follow, however, in that only a few of those patients whose behavior pattern involves dangerous assault generate a 4′3 profile. In order to be able to predict all instances of dangerously assaultive behavior we must continue to look for additional test patterns (P_2, P_3, etc.); and it is conceivable that some of these patterns may be obtained most efficiently with tests other than the MMPI.

In conclusion, Holt (1958) may have been quite right in stating that the beleaguered clinician feels he has "been sold up the river" in discussions of actuarial versus clinical prediction. This may, however, be the result of his use of inappropriate strategy in testing his observations regarding the validity of some of his psychometric instruments. This unfortunate and frustrating state of affairs need not continue.

The increasing availability of high-speed data-processing equipment makes possible the critical yet relevant examination of the predictive and concurrent validity of a wide variety of highly specific patterns of responses to any of our psychometric instruments, provided those patterns can be reduced to digital form. There are several such actuarial programs underway at the present time using the MMPI (Gilberstadt & Duker, 1961; Kleinmuntz, 1963; Marks & Seeman, 1963; Pearson, 1961). In principle, however, there is no reason why the validity of a number of other testing instruments cannot be appropriately evaluated. There is the possibility at least that there may be specific patterns of Rorschach responses (or even patterns of scorable DAP performances) which are predictive of some real-life behavior.

REFERENCES

AINSWORTH, MARY D. Problems of validation. In B. Klopfer, Mary D. Ainsworth, W. G. Klopfer, & E. R. Holt (Eds.), *Developments in the Rorschach technique.* Vol. 1. Yonkers-on-Hudson, N.Y.: World Book, 1954.

CRONBACH, L. J., & GLESER, GOLDINE. Assessing similarity between profiles. *Psychol. Bull.,* 1953, **50,** 456–473.

GILBERSTADT, H., & DUKER, JAN. Case history correlates of three MMPI profile types. *J. consult. Psychol.,* 1960, **24,** 361–367.

GILBERSTADT, H. A modal MMPI profile type in neurodermatitis. *Psychosom. Med.,* 1962, **24,** 471–476.

HOFFMAN, P. J. The paramorphic representation of clinical judgment. *Psychol. Bull.,* 1960, **57,** 116–131.

HOLT, R. W. Some statistical problems in clinical research. *Educ. Psychol. Measmnt.,* 1950, **10,** 609–627.

HOLT, R. W. Clinical and statistical prediction: A reformulation and some new data. *J. abnorm. soc Psychol.,* 1958, **56,** 1–12.

KLEINMUNTZ, B. Personality test interpretation by digital computer. *Science,* 1963, **142,** 416–418.

MARKS, P. A., & SEEMAN, W. *The actuarial description of abnormal personality.* Baltimore: Williams & Wilkins, 1963.

MEEHL, P. E. *Clinical* vs. *statistical prediction.* Minneapolis: Univer. Minnesota Press, 1954.

MEEHL, P. E. A comparison of clinicians with five statistical methods of identifying MMPI profiles. *J. counsel. Psychol.*, 1959, **6**, 102–109.

MEEHL, P. E., & DAHLSTROM, W. G. Objective configural rules for discriminating psychotic from neurotic MMPI profiles. *J. consult. Psychol.*, 1960, **24**, 375–387.

MOSEL, J. N., & ROBERTS, JANE. The comparability of measures of profile similarity: An empirical study. *J. consult. Psychol.*, 1954, **18**, 61–66.

PEARSON, J. S. Automation techniques in personality assessment. Proceedings of the Mayo Clinic, 1962, **37**, No. 3, 61–82.

(Received December 17, 1963)

Psychological Review
1964, Vol. 71, No. 6, 524–525

COMMENTS ON "A RESONANCE THEORY OF 'MICROVIBRATIONS'"

HUBERT ROHRACHER

University of Vienna, Austria

Experiments pertaining to the theoretical interpretation of "microvibration" (permanent micromovements of the human body) are quoted which exclude any resonance effects and thus are contradictory to Williams' hypothesis that the frequency of microvibration originates in resonance effects.

In his paper "A Resonance Theory of 'Microvibation'" Williams (1963) tries to show,

> that the periodic microvibration of the body surface recorded by him [Rohracher] does not require the assumption that there exists any physiological event occurring at that frequency [p. 547].

Williams proposes a "resonance theory" suggesting "that the frequency is determined by the physical characteristics of the body-transducer system used for its recording [p. 547]." In an attempt to validate his theory Williams reports a frequency comparison of records published in my 1949 monograph (Rohracher, 1949), which were obtained with two different types of transducers (piezoelectric and electrodynamic pickups). He finds differences between these pickup types (the piezoelectric pickup yielding frequencies in the 6.5–8.5 cps range, the electrodynamic one in the 9–13 cps range), but does not take into account that the two types of transducer differ also with regard to the size of the body surface they are applied to: The piezoelectric pickup has only point-by-point contact with the body surface (through a needle which is inserted into a phonograph pickup) whereas the electrodynamic pickup manufactured by Philips has a relatively large contact surface (5 millimeters in diameter) thus covering many more muscle units than the needle. Moreover, the records examined by Williams come from different subjects.

The differences in frequency—so Williams asserts—suggest that microvibration frequency should be regarded an artificial effect of the transducer-body system, "determined by the physical characteristics of both [p. 552]." This hypothesis can be tested by using methods of recording which do not involve the attachment of any load to the part of body from which the record is taken. In such a case there cannot be any resonance because nothing is present which were to resonate. I have carried out experiments of this type using two different procedures. Of these Williams refers to only one: the direct microscopic method without any electrical transducer. (A diaphragm is placed on the subject's forearm. A light beam passing through the diaphragm is recorded and reflects the movements of the diaphragm). However, Williams does not report the weight of the diaphragm: this was only 0.6 gram. It is difficult to imagine how as little a weight should cause any significant resonance effect. The second experiment without any electrical transducer (published in my 1949 monograph, together with sample records and figures showing the experimental setup) is not mentioned in Williams' report. In this latter case the subject's arm rests—without any load—between the two plates of a large condenser: micromovements of the arm change the capacity of the oscillating circuit which the condenser is part of. Both methods yielded frequencies identical to those obtained when using an electrodynamic or a piezoelectric pickup.

This would leave us only with the possibility that—even without any load being attached to it—the subject's arm itself represents a resonance system which is activated by the pulse and thus produces the microvibration. However, this hypothesis has been rejected by Sugano (Sugano & Inanaga, 1960) in experiments in animals: if the motor innervation of an extremity is interrupted microvibration of this extremity ceases almost completely although blood circulation and pulse remain unchanged.

On the basis of these experiments I have no doubt that the frequency of microvibration does represent a real parameter of an organismic process which takes place, permanently, in the musculature (alternating contractions of single motor units).

In recent years the author did not use the above mentioned two methods any more as they require a very complicated experimental setup. (When using the direct microscopic recording technique it is difficult to focus the light beam on the diaphragm. The capacitive method is unsuitable for comparative amplitude measurements because it is impossible to standardize the distance between the subject's arm and the condenser plates—be it for the simple reason that there are differences between subjects in the size of the arm.) Therefore the author came to use electrodynamic and piezoelectric transducers exclusively in his recent experimental work. At present the author works with a very small seismic piezoelectric pickup weighing as little as 15 grams which has been developed by Japanese investigators.

The pickups used in my own experiments are required to have no resonance effect whatsoever in the microvibration frequency range. Therefore, for many years, all pickups used in my laboratory have been checked for their precise resonance frequencies. (A calibration apparatus consisting of a vibrating table which is driven by a low frequency generator has been constructed specifically for this purpose.) In measuring microvibration I am using only transducers with a resonance frequency which is either damped perfectly or lies in a frequency range far outside that of microvibration.

REFERENCES

MARKO, A. R. Optisch-mikroskopische Registrierung der Mikrovibration des menschliken Körpers. *Mikroskopie*, 1959, **14**, 102–105.

ROHRACHER, H. Mechanische Mikroschwingungen des menschlichen Körpers. Wien: Urban & Schwarzenberg, 1949.

ROHRACHER, H. Ständige Muskelaktivität ("Mikrovibration"), Tonus und Konstanz der Körpertemperatur. *Z. Biol.*, 1959, **111**, 38–53.

ROHRACHER, H. Permanente rhythmische Mikrobewegungen des Warmblüterorganismus ('Mikrovibration'), *Die Naturwissensch.*, 7, 1962, 145–150.

SUGANO, H., & INANAGA, K., Studies on minor tremor, Reprinted from *Jap. J. Physiol.*, **10**, 1960, 246–257.

(Received April 27, 1964)

Psychological Review
1964, Vol. 71, No. 6, 526-527

A RESONANCE THEORY OF "MICROVIBRATIONS":

A REPLY TO ROHRACHER

JAMES G. L. WILLIAMS

Nebraska Psychiatric Institute, University of Nebraska College of Medicine

It is suggested that the resonance theory adequately accounts for those of Rohracher's findings which he quotes as being inconsistent with it and that there is at present no evidence for or need to suggest a relationship between the frequency of microvibrations and that of the "alternating contractions of single-motor units." This suggestion in no way detracts from the importance or value of the microvibration phenomenon itself.

Rohracher criticizes the resonance theory and supports his own contention that "the frequency of microvibrations does represent a real parameter of an organismic process which takes place, permanently, in the musculature (alternating contractions of single motor units)" by disagreeing with what he believes to be two critical points in the development of my argument (Williams, 1963).

First, he claims that,

> In *an attempt to validate his theory* (my italics) Williams reports a frequency comparison of records published in my 1949 monograph . . . which were obtained with two different types of transducer. . . . He finds differences between these pickup types . . . but does not take into account that the types of transducer differ also with regard to the size of the body surface to which they are applied [p. 524].

This is incorrect: the limitations of the kind of retrospective analysis made on data obtained by scoring published reproductions of records the descriptions of which were not always definitive were most clearly and explicitly emphasized. Certainly no attempt to *validate* a theory was made: it was concluded only that the analysis did "tend to support the hypothesis that the frequency is dependent on the transducer assembly used [p. 551]." (Ironically, Rohracher himself now adds weight to this conclusion by ascribing the apparent frequency differences to differences in the dimensions of the probes of the two types of transducer.) Later, however, I did describe—and here it *was* as part of the validation of the resonance theory—a critical experiment especially designed to overcome the acknowledged limitations of this retrospective analysis by comparing the frequencies obtained from two transducer assemblies differing only in their resonant frequencies and using the same probe on the same site with the same subjects. The differences in the recorded frequencies were highly significant ($p < .001$, one-tailed test) and it was pointed out that "This result accords well with the previous re-analysis of Rohracher's data [p. 554]." Rohracher makes no reference to this experiment.

Secondly, Rohracher describes two experiments the results of which he suggests are incompatible with the resonance theory and points out that in the 1963 paper, I refer specifically only to the light-beam experiment and not to the capacitative one. Actually, the former was cited merely as being typical of those of his "experiments which do not make use of electromechanical transducers to record the vibrations [p. 552]." The use of the resonance theory to explain the results of this particular experiment was offered as a specific illustration of its general applicability. The capacitative experiment would have served equally well to illustrate the point: in this case the mass, spring and damping constants of the capacitative pickup (as compared with only the last two in the light-beam experiment) would all be equated with zero and only those associated with the arm itself would determine the natural frequency of the microvibrations. The resonance theory does *not* predict that the presence or absence of any size of load would in itself determine the occurrence or nonoccurrence of microvibrations (as, indeed, is confirmed by Rohracher's own experiments). However, it is extremely important to note that if a high damping constant is associated with aperiodic movements which occur at sufficiently widely spaced intervals, then, according to the resonance theory, each oscillation will die away relatively abruptly and a recording will show intervals containing no apparent activity. As a matter of fact, Rohracher's capacitative experiment *was* replicated, had previously been reported (Williams, 1956) and in this instance gave recordings in which "only aperiodic changes in the level of the base line were found," suggesting that a high damping constant may have been involved. Rohracher's obtaining a microvibration wave form with a frequency which was "identical to those obtained when using an electrodynamic or a piezoelectric pickup [p. 525]" is inconsistent with his own current interpretation that the different pickup areas of the piezoelectric and electrodynamic transducers and, presumably, the condenser plates would all yield recordings showing *different* frequencies. Nevertheless, accepting these as identical and characteristic frequencies, it still does not follow that we are left "only with the possibility that—even without any load being attached to it—the subject's arm itself represents a resonance system which is activated by the pulse and produces the microvibration [p. 525]." And even if this were so, the subsequent statement that "this hypothesis has been rejected by Sugano (Sugano & Inanaga, 1960)" has no relevance to the acceptance or rejection of the resonance theory: in fact, the close relationship between muscle activity and microvibration amplitude was discussed in detail and illustrated in my 1963 paper and subsequently (Williams, 1964).

One final point: Rohracher concludes his criticism by observing that "The pickups used in my own experiments are required to have no resonance effect whatsoever in the microvibration frequency range [p. 525]." Once again, the precise implications of this statement were discussed at length in my 1963 article: it was pointed out that the natural frequency of the entire body-transducer system must be considered and that "this frequency will almost certainly differ from that of the transducer alone [p. 552]."

It is maintained that not only is there no need to postulate the existence of a physiological correlate of the microvibration frequency but that such a conclusion would fail to account for many of the known facts. The resonance theory appears to meet this requirement more adequately.

REFERENCES

SUGANO, H., & INANAGA, K. Studies on minor tremor. *Jap. J. Physiol.*, 1960, **10**, 246–257.

WILLIAMS, J. G. L. A new measure of muscle tension and its application to psychiatric patients. Unpublished doctoral dissertation, University of London, 1956.

WILLIAMS, J. G. L. A resonance theory of "microvibrations." *Psychol. Rev.*, 1963, **70**, 547–558.

WILLIAMS, J. G. L. Use of a resonance technique to measure muscle activity in neurotic and schizophrenic patients. *Psychosom. Med.*, 1964, **26**, 20–28.

(Received May 22, 1964)

Psychological Review
1964, Vol. 71, No. 6, 528-530

A SUGGESTED ALTERNATIVE FORMULATION IN THE DEVELOPMENTS BY HURSCH, HAMMOND, AND HURSCH, AND BY HAMMOND, HURSCH, AND TODD[1]

LEDYARD R. TUCKER

University of Illinois

In the paper "Some methodological consideration in multiple-cue probability studies" (Hursch, Hammond, & Hursch, 1964), standardized random variables Y_e, Y_s, $X_1 \cdots X_n$ were considered such that Y_e was the distal variable in the environment, Y_s was the subject's estimate of the distal variable, and $X_1 \cdots X_n$ were the cue variables.

My suggestion involves consideration of the variables $\hat{Y}_e$ and $\hat{Y}_s$ defined by,

$$\begin{aligned} \hat{Y}_e &= \beta_{e,1}X_1 + \cdots + \beta_{e,n}X_n. \\ \hat{Y}_s &= \beta_{s,1}X_1 + \cdots + \beta_{s,n}X_n. \end{aligned} \qquad [1]$$

where the β's are multiple linear regression weights for predicting the variables Y_e and Y_s from the cue variables $X_1 \cdots X_n$. $\hat{Y}_e$ and $\hat{Y}_s$ are the predicted components of the variables Y_e and Y_s while the residuals Z_e and Z_s, defined in their Equations 4 are the errors of linear prediction of these variables. Their Equations 4 may be written,

$$\begin{aligned} Y_e &= \hat{Y}_e + Z_e. \\ Y_s &= \hat{Y}_s + Z_s. \end{aligned} \qquad [2]$$

R_e^2 and R_s^2 are the squares of the multiple correlations involved and may be determined as per their Equations 5. From the propositions of multiple linear regression theory,

$$\begin{aligned} \operatorname{var}(\hat{Y}_e) &= R_e^2. \\ \operatorname{var}(\hat{Y}_s) &= R_s^2. \end{aligned} \qquad [3]$$

as well as,

$$\begin{aligned} \operatorname{var}(Z_e) &= 1 - R_e^2. \\ \operatorname{var}(Z_s) &= 1 - R_s^2. \end{aligned} \qquad [4]$$

where my Equations 4 correspond to their Equations 6. The important point of my suggestion is that their coefficient r_a, the correlation between Y_e and Y_s can be shown to have the following relation,

$$r_a = \operatorname{Cov}(Y_eY_s) = \operatorname{Cov}(\hat{Y}_e\hat{Y}_s) + \operatorname{Cov}(Z_eZ_s). \qquad [5]$$

Or, if the correlation $(\hat{Y}_e\hat{Y}_s) = G$ and the correlation $(Z_eZ_s) = C$,

$$r_a = GR_eR_s + C\sqrt{1 - R_e^2}\sqrt{1 - R_s^2}. \qquad [6]$$

Hursch, Hammond, and Hursch developed Formula 27 (p. 55),

$$r_a = \frac{R_e^2 + R_s^2 - (\beta_{e,i} - \beta_{s,i}, r_{e,i} - r_{s,i})}{2} + C\sqrt{(1 - R_e^2)(1 - R_s^2)}.$$

where: $r_{e,i}$ = correlation (Y_eX_i), $r_{s,i}$ = correlation (Y_sX_i), and

$$(\beta_{e,i} - \beta_{s,i}, r_{e,i} - r_{s,i}) = \sum_i (\beta_{e,i} - \beta_{s,i})(r_{e,i} - r_{s,i}) = \sum d.$$

the $\sum d$ form being used in the paper by Hammond, Hursch, and Todd. The last term in the above equation and in my Formula 6 are identical. A difference occurs in the form of the first term on the right sides of these equations. These two forms are equal, however,

$$GR_eR_s = \frac{R_e^2 + R_s^2 - \sum d}{2}. \qquad [7]$$

---

[1] The two articles to which the present paper refers are "Some methodological considerations in multiple-cue probability studies" by Hursch, Hammond, and Hursch and "Analyzing the components of clinical inference" by Hammond, Hursch, and Todd.

These two forms of the term may be differentially useful for different purposes.

The form of Equations 5 and 6 has an interesting parallelism in the two terms on their right side. Each of these terms involves a covariance between components: the first term involves the linearly estimated components of the distal variable and the subject's estimate of the distal variable, while the second term involves the residuals in these linear estimates. In a geometric analogy, the distal variable may be represented as a vector which is the sum of two orthogonal vectors. The linearly estimated component is a vector in the space of vectors representing the cue variables and the residual component is a vector orthogonal to this space of the cue vectors. The subject's response is represented by a similar vector which is the sum of two similar component vectors. R_e is the length of the vector for the linearly estimated component of the distal variable and R_s is the length of the vector for the linearly estimated component of the subject's response vector. G is the cosine of the angle between these two linearly estimated component vectors. The lengths of the two residual vectors are $\sqrt{1 - R_e^2}$, and $\sqrt{1 - R_s^2}$ with the cosine of the angle between these two vectors being C.

The direction of the vector for the linearly estimated component of the subject's responses is dependent on the relative values of the regression weights of the subject's responses on the cue variables. The length of the vector is dependent on the linear predictability of the subject's responses. G is a measure of the linearly predicted component of the subject's response variable in relation to the linearly predicted component of the distal variable and may be considered separately from the length measure, R_s, of the linearly predicted component of the subject's response variable.

Interpretive value of the correlation G between the linearly predicted components can be illustrated on the data from the study by Grebstein (1963) on estimation of IQ from Rorschach psychograms involving 10 cues. The IQs of 40 individuals were estimated by each of 15 clinicians. Five of the clinicians were graduate students and were termed "Naive," 5 more of the clinicians were second-year Veterans Administration trainees and were termed "Semisophisticated," and the remaining 5 clinicians were professional clinical psychologists and were termed "Sophisticated." The values of G for each of these clinicians and the median value of G for the three groups are given in the following table.

TABLE 1

	Values of *G*		
	Naive	Semisophisticated	Sophisticated
Median	.526	.716	.763
Judge 1	−.244	.593	.763
Judge 2	.401	.700	.727
Judge 3	.545	.716	.669
Judge 4	.786	.799	.874
Judge 5	.526	.829	.831

The interesting trend in correlation size from the Naive group to the Semisophisticated group to the Sophisticated group is statistically significant only to the extent that the correlations for the Naive group are below the correlations for the other two groups. These results could indicate a possible objective in training clinicians in evaluation of Rorschach psychograms for prediction of IQ to obtain like directedness of their linearly estimated components to the multiple linear regression predictions. This can be obtained by proportionate matching their regression weights to the multiple linear regression weights. The appropriate length of the residual components should be determined by the correlation of the trainees' residual components with the residual components of IQ. A second objective of training could be to increase these latter correlations; however, as noted by Hammond, Hursch, and Todd, the possibilities of this case should be the subject of research.

REFERENCES

GREBSTEIN, L. Relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment test. *J. co[illegible]. Psychol.*, 1963, **37**, 127–132.

HAMMOND, K. R., HURSCH, CAROLYN J., & TODD, F. J. Analyzing the components of clinical inference. *Psychol. Rev.*, 1964, **71**, 438–456.

HURSCH, CAROLYN J., HAMMOND, K. R., & HURSCH, J. L. Some methodological considerations in multiple-cue probability studies. *Psychol. Rev.*, 1964, **71**, 42–60.

(Received July 13, 1964)